THE ANALYSIS OF AN ENZYME (CelA) AND A GENE SYSTEM (*abg*) INVOLVED IN THE UTILIZATION OF LIGNOCELLULOSE IN THE RUMEN

by

Gordon Douglas Brown

A thesis presented for the degree of

DOCTOR OF PHILOSOPHY

in the Department of Microbiology

UNIVERSITY OF CAPE TOWN

SOUTH AFRICA

January, 1996

CERTIFICATION OF SUPERVISOR

In terms of paragraph GP 8 of "General Rules for the degree of Doctor of Philosophy (PhD)" I, supervisor of the candidate Gordon D. Brown, certify that I approve of the incorporation in this thesis of material that has already been published or submitted for publication.

Professor J. A. Thomson
Head of Department
Department of Microbiology
University of Cape Town

ABSTRACT

As lignocellulose represents an abundant renewable resource, research is in progress to obtain a better understanding of the natural mechanisms whereby this resource is utilised. Of particular interest is the degradation of forage in the rumen and one research goal is to ultimately increase animal productivity through an improvement in lignocellulose utilisation. However, although the mechanisms behind lignocellulose utilisation are reasonably well understood, relatively little is known about the mechanisms which occur in the rumen. Thus, the aim of this thesis was to gain more insight into the mechanisms of lignocellulose utilisation which occur in the rumen.

Initially this research was focused on the poorly characterised exo-acting cellulases from rumen bacteria. Preliminary enzymology studies on one cellulase from *Ruminococcus flavefaciens* FD-1, previously isolated in this laboratory, indicated that an exo-acting cellodextrinase, CelA, had been isolated. In this report, the enzyme was purified and biochemically characterised and was shown to be an exo-acting cellodextrinase.

Optimal conditions for this enzyme were similar to those found in the rumen, namely pH 6.5 and 41°C. The enzyme was inhibited by transition metals, suggesting sulfhydryl groups in the active site, and cellobiose, but did not require reducing conditions for activity. The enzyme was highly active on cellodextrins and its hydrolysis products indicated that it was operating in an exo-acting fashion. However, the enzyme also possessed weak endoglucanase, cellobiosidase and xylanase activities.

A search for another exo-acting cellulase in a gene bank of *R. flavefaciens* was unsuccessful. As *Clostridium longisporum* was capable of utilising lignocellulosic substrates, yet was relatively uncharacterised, we decided to search for exo-acting cellulases from this organism. A number of clones were isolated which, although possessing exo-acting cellulase type qualities on agar plates, were unable to be characterised in liquid assays. Sequence analysis

revealed that an aryl-ß-glucoside operon had been isolated. As this was the first such system found in a Gram positive organism, and the first genes which link the utilisation of phenolic-containing groups to lignocellulose digestion in the rumen, it was analysed further.

The aryl-ß-glucoside (*abg*) system from C. *longisporum* consisted of three genes (*abgG*, *abgF* and *abgA*) and was very similar to aryl-ß-glucoside systems from the Gram negative *Enterobacteriaceae*. The *abgF* gene encodes a phosphotransferase dependent enzyme II, specific for the transmembrane transport of aryl-ß-glucosides, and the *abgA* gene encodes a phospho-ß-glucosidase, which is active on the intracellular phosphorylated forms of salicin and arbutin. Although the protein encoded by *abgG* is homologous to a number of regulatory proteins, including those of the *Enterobacteriaceae* systems, the production and function of this protein was not demonstrated in the heterologous *E. coli* host. As the *abg* system contained a number of conserved regulatory elements, termed the Box A/ Box B motifs and a 5' promoter, it appeared to be functioning as an operon. C. *longisporum* was shown to be able to utilise aryl-ß-glucosides for growth.

A number of other C. *longisporum* genes were isolated in these experiments and, as C. *longisporum* is genetically uncharacterised, were also examined. The proteins encoded by these genes included a truncated methyl-accepting chemotaxis protein (*macA*), a T-box tryptophanyl tRNA synthetase (*trsA*), a P_{II} protein (*glnB*), and one gene of unknown function. Although not central to this thesis the proteins predicted from the nucleotide sequence of each of these genes were examined in some detail and compared to proteins to which they are similar.

ACKNOWLEDGEMENTS

It is never easy to express appreciation to people who have taught, helped, cajoled and supported you. I have received all of these things and am indebted to all who provided them. Professor Jennifer A. Thomson bestowed on me her support and her knowledge and granted me the most important of all things, the freedom to develop. Thus, it is to her that I wish to bestow my greatest thanks.

My lab mates in 204, who have put up with my misbehavior and pranks over the years, deserve my thanks and gratitude. Katrina, Caroline, Tichaona, Jacquie, Paul and Karin .... a big thank you for you support and comraderie, not to mention all the cakes and chocolates. Paul, further thanks for your advice and for the standards you set, from which I tried to learn and to emulate.

The department was a dynamic interacting cohesive group of people who provided an energising and motivating environment, which I deeply appreciated. Charlie, Keyam and Stanley, thank you for your help in keeping the lab well supplied. Ann, Di, Anne-Marie, Genevieve, Nikki and Alet, thank you for all your help in acquiring, organising and advising. Will, Cliff, Ant, Michael, Joe, Jono, Brendan, Ann, Neil, Katrina, Ros, Shelly, Andy, Sean, Declan and the rest of the crew (past and present) thanks to you for your friendship, advice and support. You made me smile, you made me laugh and you made me, on occasion, inebriated.

I am indebted to Will Bourn and Professor Douglas Rawlings for the insights, knowledge and help which they imparted to me. Also, a big thank you to Dr. Dennis Maeder, Helen Collet, Ros Powles, Di James, Mauven Graham and Mohammed Jaffer for all their assistance and their patience with me. I would also like to thank Dr. Barry Hall, Dr. Greg Blatch and Dr. Barbara Bachmann for the various *E. coli* strains they generously gave to me. Thanks also to Kenneth Palmer for proofreading parts of this manuscript.

To my parents and family, I have no means of expressing my thanks and gratitude. Their love, support and faith in me was limitless.

Janet, my wife to be, your smile was always enough. F^2N^2D.

Finally I would like to acknowledge the Foundation for Research Development (FRD) and the University of Cape Town for financial support.

CONTENTS

ABBREVIATIONS

A	adenine
aa	amino acid
ABC	ATP binding cassette
Abs	absorbance
amp	ampicillin
AMP	adenosine monophosphate
ATP	adenosine triphosphate
ATPase	adenosine triphosphatases
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine
cAMP	cyclic AMP
CBD	cellulose binding domain
CFU	colony forming units
clm	choloramphenicol
CMC	carboxymethyl cellulose
CMCase	carboxymethyl cellulase
DAS	double antibody sandwich (ELISA)
DIG	digoxigenin
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleicacid
DNase	deoxyribonuclease
DNS	dinitrosalicylic acid
DTT	dithiothreitol
EC	Enzyme Commission
EDTA	ethylene diamine tetra-acetic acid
EI	enzyme I
EII	enzyme II
ELISA	enzyme linked immunosorbent assay
EMP	Embden-Meyerhof-Parnas (pathway)
EtBr	ethidium bromide
g	standard gravitational acceleration
g	gram
G	guanine
GCG	Genetics Computer Group
GS	glutamine synthetase
hr	hour(s)
HPLC	high performance liquid chromatography
HPr	histidine protein
Ig	immunoglobulin
IPTG	isopropyl-thio-ß-galactoside
IS	insertion sequence
IU	international unit

Abbreviations cont.

kb	kilobase
kDa	kiloDaltons
l	litre
LB	Luria-Bertani
min	minutes
MES	4-morpholinoethanesulfonic acid
MCP	methyl-accepting chemotaxis protein
MOPS	3-[*N*-morpholino]propanesulfonic acid
MUC	methylumbelliferyl cellobiopyranoside
MUCase	methylumbelliferyl cellobiopyranosidase
MUG	methylumbelliferyl glucoside
MUX	methylumbelliferyl xyloside
NADP	nicotinamide adenine dinucleotide phosphate
NBT	nitroblue tetrazolium chloride
NCBI	National Centre for Biotechnology Information
NRF	non-rumen fluid (medium)
nm	nanometer
OD	absorbance
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PC	phosphate-citrate (buffer)
PBS	phosphate buffered saline
PEP	phosphoenolpyruvate
pp	pages
pNPC	p-nitrophenol-ß-D-cellobiopyranoside
pNPCase	p-nitrophenol-ß-D-cellobiopyranosidase
pNPG	p-nitrophenol-ß-D-glucoside
pNPX	p-nitrophenol-ß-D-xyloside
PTS	phosphoenolpyruvate-dependent phosphotransferase system
r	resistance
RNA	ribonucleic acid
RNase	ribonuclease
s	seconds
SDS	sodiumdodecylsulphate
T	thymine
tet	tetracycline
Tris	tris(hydroxymethyl)aminoethane
tRNA	transfer RNA
TSE	tris sucrose EDTA (buffer)
UV	ultraviolet
v/v	volume/volume ratio
w/v	weight/volume ratio
YT	Yeast Tryptone

ONE- AND THREE- LETTER CODES USED FOR AMINO ACIDS

Amino acid	Three letter code	One letter code
Alanine	ALA	A
Arginine	ARG	R
Asparagine	ASN	N
Aspartic acid	ASP	D
Cysteine	CYS	C
Glutamine	GLN	Q
Glutamic acid	GLU	E
Glycine	GLY	G
Histidine	HIS	H
Isoleucine	ILE	I
Leucine	LEU	L
Lysine	LYS	K
Methionine	MET	M
Phenylalanine	PHE	F
Proline	PRO	P
Serine	SER	S
Threonine	THR	T
Tryptophan	TRP	W
Tyrosine	TYR	Y
Valine	VAL	V

Chapter 1

The production and utilization of soluble lignocellulosic substrates in the rumen

1.1 General Introduction

Cellulose is the most abundant organic macromolecule on earth. The immense amount of carbon converted by plant photosynthesis to cellulose and other organic compounds influences the biospheric carbon cycle, as the release of carbon through the degradation of these compounds is rate limiting (236). This limitation results from the highly recalcitrant forms these molecules acquire in plants.

Cellulose, as a molecule, is not resistant to degradation. The polysaccharide is a highly elongated polymer of glucose monomers covalently linked through ß-1,4 linkages and is stabilised by intramolecular hydrogen bonds. As each successive glucose unit is rotated 180° the basic unit of cellulose is the disaccharide, cellobiose. The ability to form intermolecular hydrogen bonds allows cellulose polymers to form parallel arrays called microfibrils. The variation in order of these microfibrils, from highly ordered (crystalline) regions to less ordered (amorphous) regions, influences the susceptibility of the microfibrils to degradation.

The degradation of cellulose is further limited when it is complexed with other molecules. In the cell walls of higher plants, cellulose fibers are embedded in an amorphous matrix whose composition varies within and between plant species. The complexed nature of these molecules gives plant cells the strength to cope with turgor pressure and to withstand mechanical stress (5).

In secondary plant cell walls this matrix is composed predominantly of hemicellulose and lignin which, together with cellulose, is termed lignocellulose. Hemicellulose is a heterogeneous group of complex branched carbohydrate polymers whose composition varies depending on the plant species and stage of development but includes glucans, mannans and xylans. Xylans are a major component of woody tissue and are second only to cellulose for their abundance in nature. Lignin is a complex polyphenolic macromolecule randomly constructed

through the twenty different types of linkages occurring between the three major precursors, *p*-coumaryl, coniferyl and sinapyl alcohols. The ability of cellulose, hemicellulose and lignin to form covalent and non-covalent linkages produces the strong complex lignocellulosic matrix used by higher plants (5,37,208).

The degradation of these abundant complex structures is carried out by microorganisms. Having evolved complex arrays of degradative enzymes these organisms can utilise lignocellulosic materials as substrates for growth. As a consequence, the organisms are exploited both in nature and by man. To man, plant biomass represents a renewable resource which can be used to produce a variety of products including food, feed, fuels and chemicals. Knowledge of the mechanisms involved in the degradation and utilisation of the complex and recalcitrant lignocellulose is therefore, aside from the inherent interest, economically invaluable.

1.2 The rumen and its microorganisms

Although there are numerous environments in which organisms capable of utilising lignocellulosic material can be found, none can be more interesting or invaluable than the rumen environment and its organisms. The evolution of a rumen and symbiotic relationship with lignocellulosic degrading microorganisms has allowed ruminants to use plant biomass as feed. By utilising this biomass the microorganisms provide a source of energy and protein for the animals which, in turn, have provided a source of food and clothing for man.

1.2.1 The rumen

Ruminants have evolved a complex set of physiological and anatomical adaptations allowing them to maintain a population of microorganisms which balances both the needs of the animal and the amount of microbes needed to metabolise the resistant lignocellulosic components of the plant biomass. The muscular stomach of these animals is composed of four chambers; the rumen,

reticulum, omasum and the abomasum. The rumen and reticulum act as the fermentation vat where microorganisms degrade and metabolise the forage components. These components are regurgitated and remasticated, during rumination, before eventually passing with microbial biomass into the omasum. Here water and fermentation products are absorbed. The remaining products and microbial biomass are then passed into the abomasum where digestive enzymes, similar to those of monogastric animals, degrade this biomass to provide a source of protein and other nutrients for the animal (52,109).

The fermentation of forage components in the reticulum and rumen produces volatile butyric, acetic and propionic acids. These acids along with amino acids and some peptides are absorbed into the bloodstream from the rumen fluid through the large papillae lining the rumen and provide energy-containing compounds for the animal. As the production of these acids negatively affects microbial growth, ruminants maintain a pH level of 6.0-6.7 through the production of large amounts of alkaline saliva which is secreted into the mouth during feeding and other periods. The interactions of the fermentation acids with bicarbonate components of the alkaline saliva produces the carbon dioxide that is present in the rumen (109).

The volatile fatty acids produced are generally not utilised further by these microorganisms. More energy is obtained from the production of these compounds, through the metabolism of carbohydrates, than would be obtained from further metabolism of these acids themselves. Subsequently, the microbial growth rate is faster on lignocellulose than would be obtained if the organisms metabolised these acids to methane and carbon dioxide. The methane present in the rumen is a result of the reduction of carbon dioxide with H_2 by certain methanogenic bacteria (109,169).

The carbon dioxide, methane and other gases make the rumen anaerobic. Any free oxygen is rapidly removed through the actions of a small number of rumen organisms which are capable of utilizing O_2. In anaerobic environments

carbohydrates, such as those found in plant biomass, represent an ideal source of energy for organisms to obtain energy through substrate level phosphorylation (109).

1.2.2 The rumen microbes

A great diversity of microbes including fungi, protozoa and bacteria, are utilised in the digestive system of ruminants. Although not all microbes are capable of individually degrading lignocellulosic materials they each have a role to play in their niche of the rumen microcosm. The growth of organisms not capable of degrading these materials themselves is supported by their ability to compete for the degradation products (226).

Until the discovery of strict fungal anaerobes in the rumen, fungi were regarded as being either aerobic or facultative anaerobes. These anaerobic fungi, such as the anaerobic chytridiomycetes, are present in the rumen in low numbers but are able to ferment most of the major polysaccharides found in plants. Rumen fungi probably play an important role in the digestion of fibre as they have an ability to access polysaccharides, through mechanical or enzymatic action, in the plant cell walls which is not available to the other rumen microbes (23,37).

Rumen protozoa, including holotrich and entodiniomorphid ciliates, have a role in the rumen which is not clearly defined. Although they are not absolutely required for the growth and development of the ruminant, these microorganisms influence the rumen ecosystem and are responsible for one-quarter to one-third of the fibre degraded. The ability of these protozoa to ingest and degrade rumen bacteria, influences the proportion of energy containing compounds and protein available to the host animal (4,222).

Their numbers, diversity and ability to degrade fibrous lignocellulosic material make bacteria the most important group of microorganisms in the

rumen. They appear to be responsible for the majority of fibre digestion and are capable of utilising most of the plant organic molecules present. The bacteria found in the rumen come from diverse backgrounds and include Gram negative and Gram positive Eubacteria and a few Archaebacteria (148,200).

Some of the predominant, and possibly the most studied, rumen bacteria are *Fibrobacter succinogenes, Ruminococcus flavefaciens* and *R. albus*. These bacteria are thought to be the most active in degrading plant cell wall compounds and are also capable of attaching onto the substrate. Although these three species are widely studied, numerous other rumen bacteria are present and many have been isolated and characterised (for a review see reference 200). It should be noted that many bacterial species present in the rumen are not capable, or are only partially capable, of degrading lignocellulosic materials (167).

The distribution of the different microorganisms in the rumen is influenced by a number of factors, of which the feed composition is probably the most important. The effect of different plant tissues having different compositions and structures favours the growth of certain microbes over others (98,168,209). In addition the distribution is also affected by many and complex interactions within and between the bacteria and other ruminal microorganisms. Some are simply a result of competition, of which an example is the inhibition of *R. flavefaciens* by *R. albus*, which outcompetes *R. flavefaciens* by producing a growth inhibiting bacteriocin (145). The study showing that the rumen bacterium *R. flavefaciens* inhibits the degradative activity, but not growth, of the anaerobic fungus *Neocallimastix frontalis*, demonstrates the more complex nature of these interactions (27).

Thus the rumen is an interesting and complex microcosm. In its role of providing a source of energy, protein and other nutrients for the host, the rumen maintains a dynamic population of microorganisms which continually ferment lignocellulosic material. A symbiotic relationship has evolved where the microorganisms provide the high energy compounds and protein for the host

animal in "return" for a controlled environment and a continuous supply of nutrients, in the form of plant biomass.

1.3 The production of soluble carbohydrates from lignocellulosic materials in the rumen.

The degradation of insoluble, recalcitrant lignocellulosic compounds to their soluble components for subsequent uptake and metabolism requires a complex set of enzymes. Degradation is affected by the composition and fine structure of the plant material and by the enzymatic capabilities of the resident microorganisms. Furthermore, as these materials are insoluble and cannot be transported into the cell their degradation must occur extracellularly.

The protection given by lignin to cellulose and hemicellulose also has an important role in limiting cell wall degradation. As the degradation of lignin requires the presence of molecular oxygen, it does not occur to a significant extent in the rumen. However, the release of some phenolic compounds, such as phenolic acids and phenolic glycosides (including salicin and arbutin), does occur and these products can be utilised by rumen microbes (37).

1.3.1 Attachment of the microorganisms to the substrate

The initial step in the degradation process is attachment of the lignocellulolytic microorganisms onto the plant substrates. The microbes colonise all of the masticated plant materials except intact, outer plant surfaces which are protected by epicuticular waxes and the cuticle. There are a number of possible mechanisms whereby microorganisms can colonise plant tissues, including cell surface-associated proteins and enzymes, adhesins, or non-specific ionic interactions. The attachment of microbes to the substrate also ensures close proximity to the released soluble carbohydrate products, which are competed for by the other rumen microbes (37,164,226).

In rumen fungi, such as *Neocallimastix*, motile zoospores are chemotactically attracted to plant tissues by the release of soluble carbohydrates from the damaged tissue. Once they have located the damaged tissues they attach and then germinate (149). Although the attachment process is poorly understood it might involve surface bound enzymes, such as those found on the conidia of the non-rumen *Trichoderma reesei* (135). The rhizoidal system of the fungal thallus then penetrates the plant tissue either by mechanical action or enzymatic process or a combination of both. As rumen fungi are proteolytic, unlike the rumen cellulolytic bacteria, they are able to penetrate proteinaceous layers, such as the primary cell wall, and gain access to the lignocellulosic secondary plant cell walls (149). These actions on plant tissues open more sites for microbial attack and therefore influence the rate of fibre digestion.

The attachment of cellulolytic ruminal bacteria to plant substrates varies between species but appears to involve protuberances on the cell-surface which have not been found in non-cellulolytic bacteria (2,3,23,201). Bacteria such as *R. flavefaciens* and *R. albus* are found loosely attached while bacteria such as *F. succinogenes* are closely associated with the substrate (37). Attachment of these organisms is thought to be mediated by surface bound enzymes and other proteins (23,59,75,201).

Although attachment to the substrate is not a major factor in determining the cellulolytic activity of a culture *per se* (23), the attachment of microorganisms is the first stage of the degradation process in the rumen and allows extensive utilization of the available substrates (37).

1.3.2 The enzymes involved and their classification

The enzymes which attack cellulose and the backbones of hemicellulose currently fall into three groups depending on their mode of action on the substrate. The endo-acting components are the most widely found types of enzymes and randomly attack internal sites of the polymers. Exo-acting

components, found more in cellulase than in hemicellulase systems, attack the polymer ends. The third group consists of enzymes which attack the small soluble substrates produced through the actions of the former enzymes. The mode of action of these enzymes determines their enzyme commission (EC) number.

Hemicellulose, however, contains a number of other molecules and requires a variety of other enzymes for its degradation. For example, as xylan is highly branched with acetyl, arabinosyl and glucuronosyl residues these side chains must be removed through the actions of appropriate enzymes such as acetylesterases, α-L-arabinofuranosidases and α-glucuronidases (208).

However, the classification of lignocellulosic enzymes simply depending on their mode of action is somewhat simplistic. Not only do lignocellulolytic organisms individually produce a variety of these enzymes but many of these enzymes tend to have overlapping substrate specificities. *F. succinogenes,* for example, has genes encoding at least nine different endoglucanases, one cellodextrinase and some hemicellulases (59,128). Recently, a revised method of assigning EC numbers has been proposed where the actual catalytic mechanism of these enzymes is used to assign them an EC number (Workshop on cellulase nomenclature, 1994). A classification scheme using hydrophobic cluster analysis (HCA) has also been developed. In this scheme, two dimensional representations of amino acid sequences are used to determine hydrophobic clusters, which are then used for sequence comparisons. Since this development, all glycosyl hydrolases have been reclassified and cellulases and xylanases have been grouped into eleven distinct families (65,72,93-95).

As the degradation of insoluble lignocellulose occurs extracellularly the lignocellulases must be secreted from the cell. In both fungi and bacteria, these enzymes may be either cell-bound or free in the medium. Furthermore, the location of these enzymes may vary depending on the stage of growth and/or the substrate utilised (24,132,135). For example, endoglucanase activity of

Fibrobacter succinogenes grown on glucose or cellobiose is primarily cell associated. However, when the cells are grown on cellulose there is a shift towards release into the medium (106,132).

1.3.3 Catalytic mechanisms

Cellulases and xylanases catalyse the hydrolysis of glycosidic bonds through acid-base catalysis (Figure 1.1). Two amino acid residues are involved, the first acting as a general acid catalyst and the second as a nucleophile. The hydrolysis of glycosidic bonds by these residues involves nucleophilic substitutions at the saturated carbon of the anomeric centre. The reaction of the nucleophile may lead to either retention or inversion of the anomeric configuration (23,71,196). It is the ability to retain or invert this configuration that will be used in the assignation of an EC number to these enzymes (Workshop on cellulase nomenclature, 1994).

The structure of the catalytic cores of some fungal and bacterial enzymes have been determined (112,166,198). Endocellulases and xylanases have open clefts which form the active site, allowing these enzymes to attack any part of the substrate polymer, whereas exocellulases have active site tunnels through which the substrate has to be threaded. Thus the main difference between exocellulases and endocellulases appears to be the mode in which the active site accesses the substrate (198).

The activity of many cellulase enzymes is affected by the amount of end-product, cellobiose, present and by the substrate itself (205). Structural studies have demonstrated that cellobiose binds in the active site and thereby acts as a competitive inhibitor (198). As previously mentioned, the features of the substrate itself, such as the degree of crystallinity and its complexity, can also affect the activity of these enzymes.

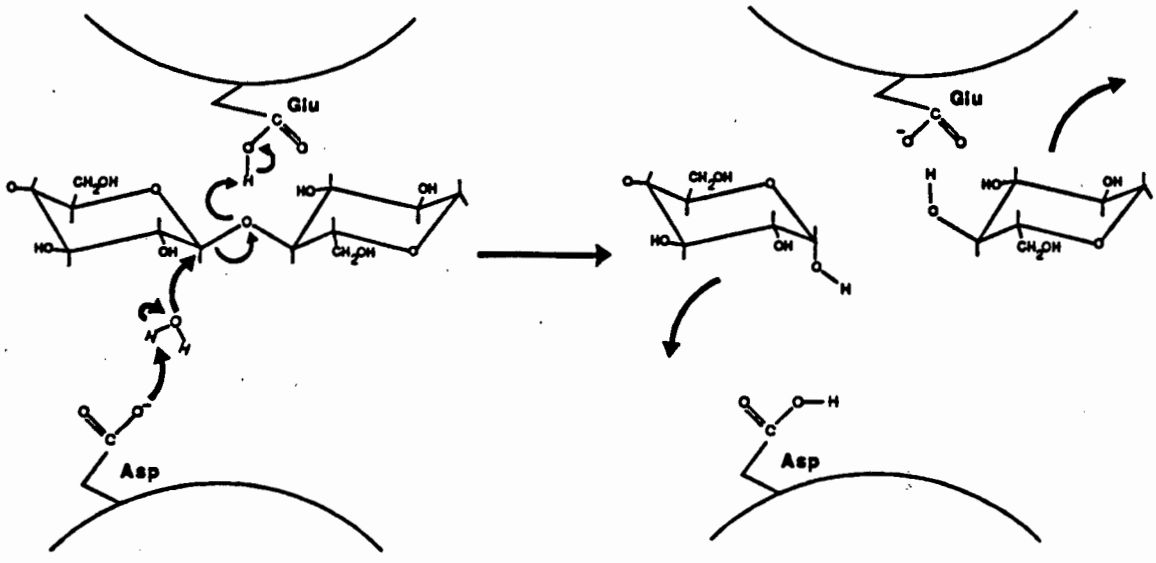


Figure 1.1 Catalysis of ß-1,4 glycosidic linkage with inversion of configuration at the anomeric carbon. The H^+ ion required for protonation of the glycosidic oxygen is provided by the Glu residue. Simultaneously, the negatively charged Asp residue promotes the ionization of a water molecule. The resulting OH^- ion interacts with the anomeric carbon, leading to a single-step nucleophilic substitution and to inversion of configuration (from (23)).

1.3.4 Enzyme domains and multimeric complexes

Besides the catalytic domain many enzymes have a number of other non-catalytic domains connected by linkers of which most are rich in proline and hydroxyamino acids. Although these linkers are thought to act as extended, flexible hinges allowing the independent function of the separate domains, their role is not always so clear. For some enzymes, spatial separation of the domains was shown not to be required for normal function (55,73,199).

Many enzymes possess discrete domains which bind to cellulose and are known as cellulose binding domains or CBDs. Although these domains are not essential for catalytic activity, they do affect the enzyme's ability to attack various substrates. Most of these domains have been grouped into four families, of which the best characterised is Type 2. This group contains only fungal CBDs and the wedge shaped structure of one has been determined (23,43,114).

The CBDs also play a role in the degradation of cellulose. Not only has the presence of a CBD been linked with the ability of cellulases to degrade crystalline cellulose, but at least one CBD actually disrupts the cellulose fibres itself by "unzipping" cellulose chains from the cellulose microfibrils. Interestingly, some xylanases and an arabinofuranosidase also possess CBDs whose role is probably to aid in the attachment of these enzymes to the plant cell wall (23,43,55,113,136).

Another group of conserved elements are the 22 amino acid repeated segments found in some bacterial cellulases and hemicellulases. The role of these segments has been clearly implicated in the aggregation of the enzymes into multimeric complexes, which are termed cellulosomes. The cellulosome of the non-rumen bacterium *Clostridium thermocellum* is the best characterised (for a review see (23)). During the early stages of growth of this organism the cellulosomes aggregate on the cell surface, becoming polycellulosomes. These structures are visualised as protuberances on the cell surface and are involved in the attachment of the bacterium to the substrate (21,23,54,71,73).

The occurrence of these aggregating systems in rumen bacteria has not been clearly demonstrated. Furthermore, the presence of repeated segments involved in the assembly of cellulosomes has not been shown for rumen enzymes. However, so-called "non-aggregating enzymes" may form associations through protein-protein interactions (23).

Although rumen bacteria have cell-surface localised cellulosome-like structures, such as *R. albus* (54), the composition of these surface protrusions is

unknown. *F. succinogenes*, for example, produces surface protrusions which contain some lignocellulases, but not in the form of cellulosomes. Rather the cellulase enzymes appear to occur individually at the bacterial surface (59,128). As discussed earlier, these protrusions are thought to play a role in attachment of the organism to the substrate.

In addition to their catalytic and non-catalytic domains many lignocellulases carry other domains whose functions are not yet clear. Some of these domains are similar to fibronectin type III domains whose function may be protein-protein interactions between other lignocellulases or they may function in anchoring the enzymes to the cell surface. Other domains resemble S-layer proteins whose function may also be to bind the enzymes to the cell surface (23,179). *F. succinogenes*, for example, produces an integral membrane protein which possesses a CBD. This protein is thought to be involved, along with other surface bound enzymes, in the tight association of the organisms to lignocellulosic substrates (59,75).

The possession of more than one catalytic domain has been described for a number of enzymes. One example is the bifunctional enzyme isolated from *R. flavefaciens* carrying both xylanase and ß(1,3-1,4)-glucanase domains (57). The role of such proteins in lignocellulose degradation is unclear, but Beguin and Aubert (1994) suggest that the bifunctional cellulase/ hemicellulase enzymes may have a role in separating the hemicellulose from the cellulose fibres.

1.3.5 Synergism in lignocellulose degradation

The efficient use of lignocellulosic substrates is only achieved by a few rumen microorganisms and the majority produce only incomplete sets of enzymes. Furthermore, only a few of these enzymes are individually capable of degrading recalcitrant substrates, such as crystalline cellulose. As a consequence, the efficient utilisation of lignocellulosic substrates depends on the synergistic interactions of all the rumen microorganisms and their enzymes (119).

One model has been derived from fungal systems to explain the synergistic interactions of non-aggregating cellulases in degrading crystalline cellulose (23). In this model, endo-cellulases act initially on amorphous regions of the crystalline fibres, producing non-reducing ends which can then be attacked by exo-acting cellulases. The cellobiose released through the actions of these enzymes is cleaved by a ß-glucosidase to glucose (Figure 1.2).

This model is, however, oversimplified and does not take into account the multiple types of cooperation reported. An example is the exo-exo synergism observed with the exocellulases of *T. reesei* (96). Furthermore, the model does not take into account the effect of structural and ultrastructural features of the substrate on these synergistic interactions (96).

Synergism has also been demonstrated for the enzymes of bacteria which use aggregating systems. A model for the degradation of cellulose by the cellulosome suggests that a scaffolding protein (CipA), which possesses a cellulose binding domain attaches to the cellulose and binds the catalytic components, mediated by the 22 amino acid repeated segments. This would bring the catalytic components in close association with the substrate and, furthermore, the orientation of the cellulosome might promote multiple cutting along a single cellulose chain (23,54,63,64,69,180,181,195).

The end products of the enzymatic attack on cellulose and hemicellulose are short, soluble carbohydrates. *F. succinogenes*, for example, degrades cellulose to glucose and cellobiose while *R. flavefaciens* degrades cellulose to cellotriose, cellobiose and small amounts of glucose (59,158). Thus, whether cell bound or secreted, a diverse range of lignocellulases degrade the cellulose and hemicellulosic components of rumen forage. Although not all the lignocellulosic components can be utilised in the rumen, the degradation of utilisable substrates is achieved by the synergistic interactions of the rumen organisms and their enzymes.

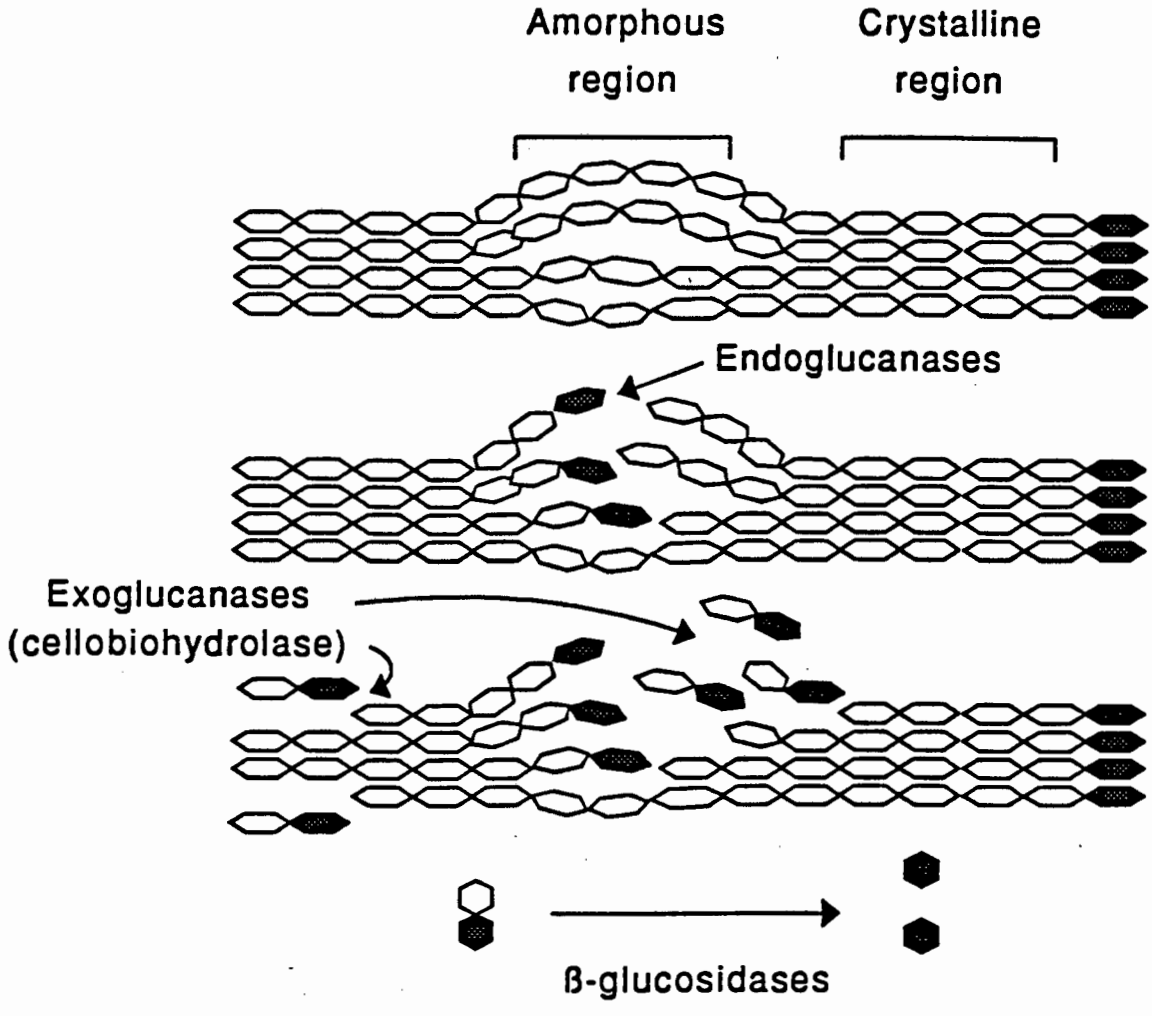


Figure 1.2 Synergism between exoglucanases, endoglucanases and ß-glucosidases. Hexagons represent glucose residues and black residues represent reducing ends (from (23)).

1.4 Utilisation of soluble carbohydrates by rumen bacteria

The soluble carbohydrates, such as glucose and cellobiose, produced through the actions of the microbial enzymes are competed for by the various microbes and are taken up and utilised for growth. As bacteria are protected by plasma membranes impermeable to carbohydrates, they require trans-membrane transport systems to bring these substrates into the cytoplasm for subsequent catabolism. Furthermore, as the concentration of extracellular soluble carbohydrates is generally low in the rumen, transport into the cell establishes concentration gradients. Thus to overcome these gradients, the transport mechanism must be energy-linked (167,169).

1.4.1 Transport across the outer and plasma membranes

Although the peptidoglycan layer of bacterial cell walls is porous and allows the diffusion of nutrients, the outer membrane of Gram negative cells does not. The permeability of this membrane is mediated by three groups of proteinaceous channels, which will only be briefly mentioned here. These groups include (i) the porins, which non-specifically mediate the passage of small, hydrophilic nutrients, (ii) the porin-like proteins which have specific binding sites and carry out facilitated diffusion of nutrients unable to rapidly penetrate the porin channels, and (iii) a high affinity outer membrane receptor and a cytoplasmic membrane-bound protein, thought to couple solute translocation and energy input, which transport large nutrient molecules (50,142,143).

Transport systems across the plasma membrane are generally classified depending on the nature of the energy source used for transport. In bacteria, there are three main groups of transport systems including osmotic shock-sensitive active transport, ion gradient-linked active transport and group translocation. Although a fourth system involving facilitated diffusion has been described, it occurs with the transport of glycerol and will not be discussed here.

Bacteria can utilise any number and combination of these systems for the uptake of the various carbohydrates (49,143).

1.4.1.1 Osmotic shock-sensitive active transport

Osmotic shock-sensitive active transport systems depend on a periplasmic binding protein, hence the sensitivity to osmotic shock, which binds the nutrient molecule. In addition to the binding protein, transmembrane transport is mediated by two integral membrane proteins and associated ATP-binding proteins. The latter function as ATPases and provide the energy for trans-membrane transport. These transporters have also been termed ATP binding cassette-type (ABC-type) transporters (49,143,206).

Although these systems occur in Gram negative bacteria, they have also been described for Gram positive bacteria (7,74,170). In the latter, which lack a periplasm, the binding protein is thought to be anchored by covalent linkage of its NH_2 terminus to lipids in the cytoplasmic membrane (74). *R. albus*, for example, may transport xylose, arabinose and glucose using this system as the transport of these carbohydrates was shown to be linked to ATP hydrolysis but not linked to group translocation or to ion gradient-linked active transport mechanisms (210).

1.4.1.2 Ion gradient-linked active transport

Ion gradient-linked active transport derives the energy required for transport from the electrochemical gradients that exist across the membrane. In this system, trans-membrane transfer, mediated by integral membrane proteins, requires the transfer of a second solute (such as H^+, Na^+ or phosphate) either in the same direction, symporter, or opposite direction, antiporter (6,143,169).

The ability to take up carbohydrates using this active transport system has been demonstrated for a number of rumen bacteria. *R. flavefaciens* utilises the

electrochemical gradient to take up cellobiose and presumably cellotriose, although the specific energy source for uptake is unknown (87). *F. succinogenes* has been shown to take up glucose and cellobiose through Na^+ symporters, while another rumen bacterium, *Selenomonas ruminantium*, takes up xylose and arabinose through pentose-proton symporters (38,59,203).

1.4.1.3 Group translocation

The mechanism of group translocation couples solute translocation with chemical modification of the transported solute. By expending metabolic energy in modifying the transported solute, the bacterium also maintains the concentration gradient by removing the substrate from the diffusion equilibrium (53). In many anaerobic and facultatively anaerobic bacteria, sugar uptake is linked with concomitant phosphorylation. The metabolic energy expended in this process stems from phosphoenolpyruvate (PEP) and the system is thus referred to as the PEP-dependent phosphotransferase system (PTS), which will be discussed below.

1.4.2 The phosphoenolpyruvate dependent: sugar phosphotransferase system

In this system, a phosphoryl group is transferred from PEP through soluble intermediates to a membrane spanning protein(s) (enzyme II) which translocates the carbohydrate with concomitant phosphorylation of the sugar (Figure 1.3). The soluble intermediates, enzyme I (EI) and histidine protein (HPr), are general PTS proteins, required for the phosphorylation of all PTS carbohydrates, while enzymes II are sugar specific PTS proteins (133).

Most enzyme I proteins consist of monomers that self-associate to form dimers. The autophosphorylation of EI (forming EI-P) in the presence of PEP

requires divalent cations and phosphorylation occurs at a conserved histidine residue (133,156). Histidine containing proteins (HPr's) are generally small heat-stable, soluble monomeric proteins which interact with both PTS and non-PTS proteins. Phosphorylation of HPr by EI-P occurs at a histidine residue within a highly conserved sequence. However, in Gram positive bacteria phosphorylation can also occur at a serine residue through the actions of ATP-dependent HPr kinases and is thought to be involved in PTS-mediated regulation, as discussed below (48,89,156,232).

The sugar specific PTS proteins, or enzymes II, have a number of autonomous domains which may be either free or fused but act in concert to translocate the carbohydrate into the cytoplasm (Figure 1.3). As a consequence, enzymes II may be composed of between one to four individual polypeptides, and a number of these proteins and their encoding genes have been isolated (156). The hydrophobic transmembrane domain, which binds and transports the sugar, and two hydrophilic domains, through which the phosphoryl group is passed to the transported carbohydrate, comprise the sugar specific transmembrane transport complex.

Due to the multiplicity of these multidomain complexes, Saier and Reizer (1992) proposed a uniform system for their classification. In this system domain IIA is the first hydrophilic domain (previously termed factor III), domain IIB is the second hydrophilic domain and domain IIC as the transmembrane domain. A fourth protein/domain found in a few PTS systems (such as the II^{man} complex of *E. coli*) acts as a second transmembrane protein and has been termed IID.

The sequence of transfer of the phosphoryl group, derived from PEP, through the intermediates has been determined. The phosphoryl group is transferred from Enzyme I to HPr and then to enzyme II domain A, which is phosphorylated at a conserved histidyl residue. The phosphoryl group is then passed through domain B, phosphorylated at a cysteinyl (or in a few cases at a histidyl) residue, to the transported sugar. The phosphorylation and

dephosphorylation of domain IIA^{glc}, in enteric bacteria, is also involved in the regulation of catabolic enzymes and other non-PTS carbohydrate permeases, as discussed below (156,174,175).

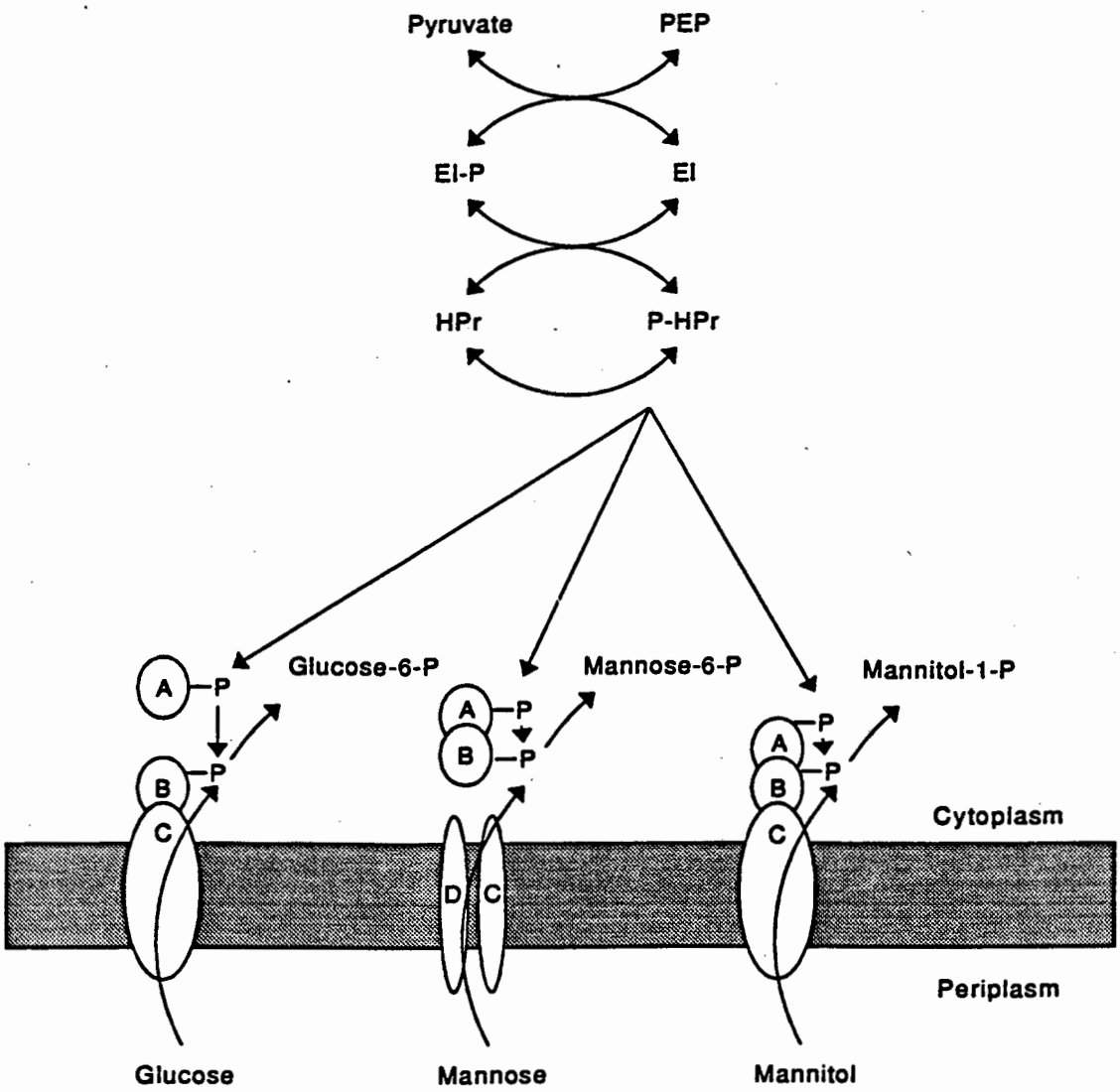


Figure 1.3 The PTS system of *E. coli*. Shown is the phosphoryl transfer from EI to the various forms of enzyme II. The filled rectangle represents the bacterial membrane (from (156)).

The mechanism of translocation of sugars into the cytoplasm by the PTS system is not clearly understood. Enzyme IIC is capable of weak translocation of the substrate into the cytoplasm by facilitated diffusion. Furthermore, the coupling between solute translocation and phosphorylation is not obligatory, although efficient uptake requires an intact complex (143,156).

Postma *et al.* (1993) have proposed a model for sugar translocation based on analysis of *E. coli* enzymes II. They propose that the sugar substrate binds with high affinity to its specific EII, which if not phosphorylated, translocates the substrate slowly by facilitated diffusion. Phosphorylation of enzyme IIB then causes a conformational change which allows rapid translocation of the substrate, with the phosphoryl group being passed to the carbohydrate. The phosphorylated carbohydrate is then released into the cytoplasm. How the substrate is actually translocated through the trans-membrane protein to the cytoplasm is unknown.

Although PTS systems have been found in a wide range of fermentative bacteria only a few have been described for rumen bacteria. Martin and Russell (1986) looked for evidence of PTS systems in six ruminal bacteria but only detected PEP dependent PTS activity in *Selenomonas ruminantium, Streptococcus bovis* and *Megasphaera elsdenii. S. bovis* was shown to phosphorylate cellobiose, glucose, sucrose and maltose by the PTS system using enzymes II specific for the uptake of these carbohydrates. *S. ruminantium,* however, had enzymes II for glucose and sucrose but maltose was cleaved extracellularly by a maltase and the products were taken up by enzyme II^{glc} (129-131).

E. coli is not normally considered a rumen bacterium but has been isolated from the rumen of young ruminants (200). Furthermore, it possess PTS dependent enzymes II for cellobiose and phenolic glycosides, such as arbutin and salicin, which are associated with lignocellulose degradation. The genes encoding cellobiose and aryl-ß-glucoside enzymes II are present in operons that are normally cryptic in laboratory strains and natural isolates of *E. coli.* However, functional genes for cellobiose utilisation have been identified in isolates from

faecal samples of some ruminants (80).

1.4.3 Metabolism of translocated carbohydrates

Once the carbohydrate has been transported into the cell it is metabolised further, providing energy for the microbial cells and resulting in the production of the volatile fatty acids which are used by the ruminant host. Glycolysis, or the Embden-Meyerhof-Parnas (EMP) pathway, is the primary pathway of hexose metabolism in rumen bacteria as it provides a maximal yield of ATP (169). Although the pentose phosphate pathway is responsible for the majority of pentose metabolism in the rumen (169), it will not be discussed here.

Once in the cell, disaccharides need to be degraded to their monosaccharide components before entering the catabolic pathways. This is carried out by hydrolases or phosphorylases which cleave the disaccharides to their monomers or to phosphorylated and unphosphorylated monomers, in the case of the latter enzymes. As phosphorylated sugars are required for subsequent catabolism, unphosphorylated monosaccharides are phosphorylated by glucokinases. Examples of these types of enzymes include the cellobiose phosphorylase of *R. flavefaciens* (15,16) and the maltose phosphorylase and sucrose-hydrolase activities described for *Streptococcus bovis* (130).

Sugars that are transported into the cell via the PTS system are already phosphorylated. Phosphorylated disaccharides are cleaved by sugar specific phospho-ß-glucosidases, such as the sucrose-phosphate hydrolase of *S. bovis* (130), to monosaccharide and monosaccharide-phosphates. The former must also be phosphorylated before entering the catabolic pathways.

To enter the EMP pathway, phosphorylated sugars must also be converted to glycolytic intermediates. These conversions are carried out by enzymes such as isomerases and mutases. Phosphoglucomutase of *S. bovis* is an example of such an enzyme, and converts glucose-1-phosphate to glucose-6-phosphate (130). Glucose-6-phosphate is then converted through a series of intermediates to phosphoenolpyruvate and then pyruvate, with the production of ATP (Figure 1.4). If the conversion of phosphoenolpyruvate to pyruvate is used to phosphorylate EI of the PTS, no ATP is formed at this step. Pyruvate is then metabolised further, with further gains in ATP, to volatile fatty acid waste products which are released from the cell (Figure 1.5).

The production of the different fatty acids in the rumen depends on the bacterial species and feed composition (109,200). *S. ruminantium*, for example, produces lactate when grown on glucose but after depletion of the hexose ferments lactate to acetate and propionate (131,168), while *F. succinogenes* ferments glucose and cellobiose to succinate and acetate (59).

The volatile fatty acids released by the microbes into the rumen are absorbed by the host animal and utilised further. The ruminant is able to obtain energy from these compounds, which may account for 50 - 60 % of the metabolizable energy intake of the animal (14).

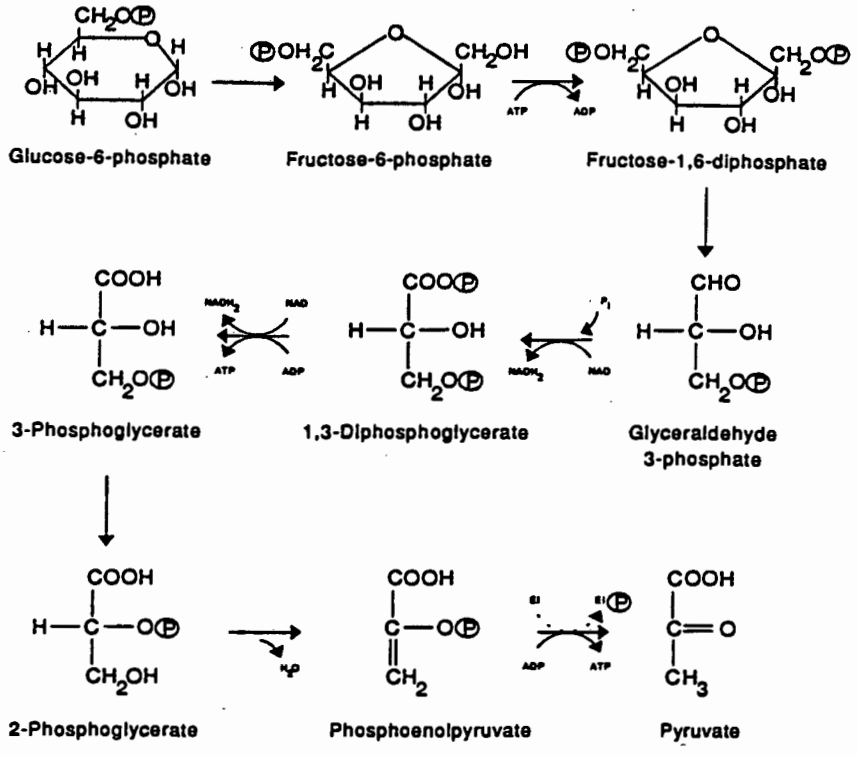


Figure 1.4 The Embden-Meyerhof-Parnas pathway of glycolysis where carbohydrates are converted to pyruvate with the generation of ATP (from (29)).

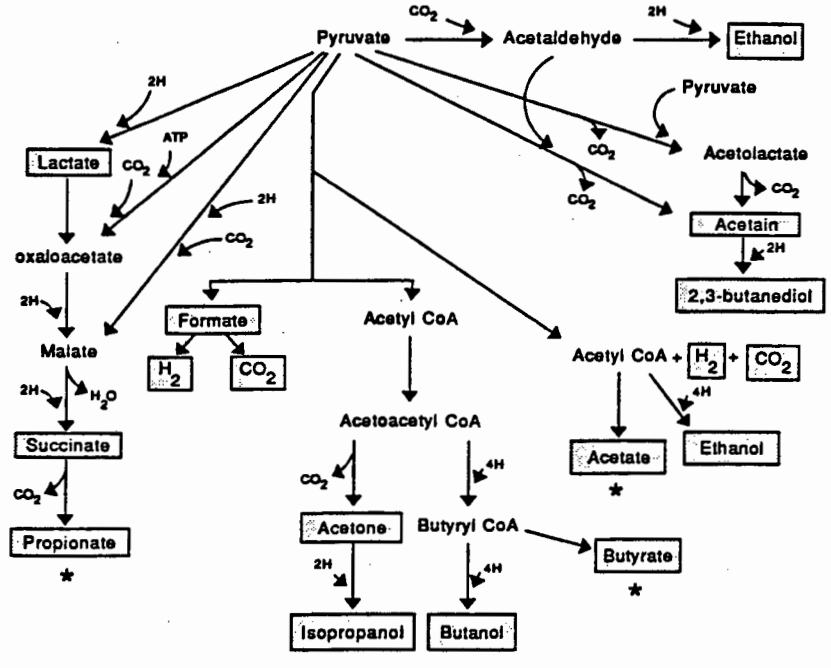


Figure 1.5 Products of pyruvate metabolism. Products enclosed in boxes accumulate during microbial fermentations, and the * indicates the principal volatile fatty acids which are formed in the rumen (from (29)).

1.5 Regulation of the production and utilization of soluble carbohydrates

Both fibrolytic and non-fibrolytic rumen microbes compete for the products of lignocellulose hydrolysis. To compete successfully, rumen microbes have developed different preferences for certain substrates over others. *R. albus*, for example, preferentially utilises cellobiose over glucose (209), while *S. bovis* utilises glucose over cellobiose (168). By having different strategies for utilising the available substrates the microorganisms are therefore able to occupy separate niches in the rumen environment (110,168). The preference in utilising of one substrate over another is broadly termed catabolite repression and is controlled by a complex variety of regulatory mechanisms (202).

1.5.1 Regulation of cellulase synthesis

The regulation of cellulase biosynthesis is poorly understood. Physiological, biochemical and mRNA analysis of cellulases has, however, suggested that two general underlying regulatory mechanisms are present in cellulolytic microorganisms, namely (i) repression of cellulase synthesis in the presence of easily metabolised low molecular weight carbon sources (catabolite repression), or (ii) induction of cellulase biosynthesis in the presence of cellulose or its degradation products, not linked to repression by low molecular weight carbon sources (23,205).

For example, although the majority of cellulases of *F. succinogenes* are synthesised constitutively, irrespective of the carbon source, the synthesis of one endoglucanase was repressed by cellobiose (107,132). Xylanase expression in *Butyrivibrio fibrisolvens* was also repressed by easily metabolisable carbon sources, but the initial production of these enzymes was induced by xylan degradation products (223). Induction of lignocellulase enzymes has also been shown for other

rumen bacteria, such as *Prevotella ruminicola* (66) and *Ruminococcus* (51,194).

A general model describing the induction of lignocellulase genes has been described (23,208). The induced genes are thought to be constitutively transcribed at low levels which is responsible for limited hydrolysis of lignocellulosic substrates, thereby producing soluble hydrolysis products. These soluble products would be then taken up by the cells and induce enzyme synthesis. The low level constitutive expression has, for example, been demonstrated for the cellulose inducible *cenB* transcript, encoding an endoglucanase of an occasional rumen isolate, *Cellulomonas fimi* (77,200). Although the true intracellular inducers (and regulatory proteins) are unknown, some, including cellobiose and sophorose, have been proposed (23,28).

1.5.2 Regulation of carbohydrate utilisation

Although cAMP and its receptor protein were once thought to be the central catabolite regulators, a number of cAMP independent regulatory mechanisms have since been identified (172,173). These catabolite repression mechanisms include utilization of other regulatory elements such as novel transcription factors (111), repressors (such as FruR or CcpA (108,115,157,174)) or protein phosphorylating enzyme systems (such as the PTS) (173). Although these mechanisms are not clearly understood it is thought that their effects are exerted at the level of transcription (173).

The regulation of lignocellulases and the control of substrate utilisation is most likely to be regulated by a number of these control mechanisms, differing between the various rumen bacteria. Furthermore, although the mechanisms of regulation in aerobic and facultative anaerobes have been well characterised, relatively little is known about regulation in anaerobic bacteria.

Recently, Cotta *et al.* (1994) compared cAMP levels in a number of ruminal

and non-ruminal anaerobic bacteria to those of *E. coli*. All but one of the bacteria studied had low levels of cAMP, suggesting that cAMP was not involved in catabolite regulation in anaerobic bacteria. In light of these results, cAMP mediated regulation will not be discussed here.

The PTS system plays a key role in a number of bacteria in regulating not only its own protein components, but also the transport systems and enzymes involved in non-PTS carbon source catabolism and the activities of the central pathways of carbon catabolism (173,178). Furthermore, although some bacteria lack PTS enzymes II they contain PTS components whose primary function must involve regulation of non-PTS functions (99,178). As PTS components do occur in ruminal bacteria, the role of the PTS in regulation will be briefly presented here. However, as most analyses of PTS mediated regulation have been made in the enteric bacteria *E. coli* and *Salmonella typhimurium* (156), the regulatory mechanisms described may or may not hold true for the rumen anaerobes possessing these systems.

Regulation of the PTS proteins occurs both at the level of enzyme synthesis and enzyme activity. Although the general PTS proteins are synthesised constitutively, the specific proteins are induced. Regulation of the specific proteins is accomplished by a number of mechanisms including repressor-operator pairs and transcriptional antitermination. Furthermore, the synthesis of these proteins is also influenced by catabolite repression (49,155,156).

Regulation of the enzyme activity of the PTS proteins is accomplished by a number of mechanisms (49,155). For the specific PTS proteins there are a number of control mechanisms which operate, including; (i) competition for binding site on the sugar permease, (ii) regulation by intracellular sugar-phosphates, (iii) regulation by the membrane potential, and (iv) competition between enzymes II for phospho-HPr.

The PTS system also regulates catabolic pathways involving non-PTS

substrates. In Gram negative enteric bacteria, the phosphorylation state of enzyme IIA^{glc} is central to this regulation. This enzyme can inhibit the activities of non-PTS carbohydrate permeases and catabolic enzymes as well as influencing adenylate cyclase activity, which produces cAMP. Although catabolite regulation mediated by enzyme IIA^{glc} has not been demonstrated in Gram positive organisms, HPr appears to have a role similar to enzyme II^{glc}. HPr, in Gram positive bacteria, can also regulate the utilisation of non-PTS carbohydrate permeases and catabolic enzymes (152,156,159,174,232).

Thus the regulation of carbohydrate utilisation is clearly complex. Although not much is known about the regulatory mechanisms in rumen bacteria, they must employ a number of different control mechanisms to utilise one substrate over another. As a consequence of this regulation, rumen bacteria are able to occupy different niches in the rumen environment.

1.6 Concluding remarks and aims of this thesis

The mechanisms behind the utilisation of recalcitrant lignocellulosic materials by rumen microbes are very intricate. The rumen and its microorganisms present a dynamic microcosm where the energy available in lignocellulose is utilised by the resident microbes, which in turn supply the animal with both protein and energy. Microorganisms, producing a variety of enzymes which degrade the lignocellulosic components, compete for the digestion products with numerous others in the rumen. Having a number of mechanisms and strategies for utilising the available carbohydrates, this competition has resulted in organisms with different substrate preferences which are controlled by complex regulatory processes.

Aside from the inherent interest, knowledge of the processes involved in lignocellulose degradation is economically invaluable. One goal is to improve

fibre digestion and metabolism in the rumen, and hence increase productivity, and research directed at the genetic modification of the rumen organisms is in progress (8,9,23). The numerous other applications of lignocellulases, from fuel production (46) to textile processing (184), further demonstrate the merits of cellulase research. The aim of the research presented in this thesis was to gain further insight into the mechanisms and processes behind lignocellulose degradation and utilisation in the rumen.

The exocellulases of rumen bacteria were analysed as they are poorly characterised and their role in lignocellulose degradation in the rumen is unclear. *R. flavefaciens* was chosen for this study as it is a predominant rumen organism and is capable of utilising lignocellulosic materials for growth. The previously isolated *R. flavefaciens* cellulase, CelA, has been cloned and sequenced (19,219,220) and was initially thought to be a endo-acting cellulase. As subsequent evidence, however, suggested that this enzyme might be an exo-acting cellodextrinase, further analysis was warranted.

R. flavefaciens produces at least two exo-acting cellulases ((67), this study), but only the sequence of CelA has been reported (19,219,220). To further characterise the exo-cellulase components of this organism an attempt to isolate novel exocellulase genes was undertaken, but was unsuccessful.

To further assess the exocellulase components of rumen bacteria, *Clostridium longisporum* was chosen for two reasons. Firstly, it has only been recently reisolated (215) and the cellulolytic capabilities of this organism have not been clearly elucidated (137,216). Secondly, it is a spore forming bacterium which may prove to be a useful attribute for future studies (215,216). Unexpectedly, a search for exocellulase expressing genes from this organism resulted in the isolation of the first PTS dependent system for the uptake and utilisation of aromatic ß-glucosides from a Gram positive organism. Due to the involvement of such systems in the utilisation of lignocellulosic degradation products, especially phenolic containing compounds, it was decided to analyse this system

in detail. The sequence and characterisation of genes involved in this system, as well as their regulation, was examined.

This work therefore examines enzymes involved in both the production and utilisation of soluble carbohydrates by rumen bacteria. Although many questions remain unanswered, it is hoped that the work presented here will answer at least some of them.

Chapter 2

Biochemical analysis of a cellodextrinase cloned from *Ruminococcus flavefaciens* FD-1

2.1 Abstract

The cellulase encoded by the previously cloned gene, *cel*A, was purified to homogeneity and characterised. The enzyme was antigenically similar to a protein of similar molecular weight in *R. flavefaciens* and was shown to be a cellodextrinase with predominantly exo-type action. CelA also displayed weak endoglucanase, cellobiosidase and xylanase activities. The enzyme had optimal activity at pH 6.5 and 41°C, was inhibited by transition metals and did not require metal ion cofactors or reducing conditions for activity. The enzyme had high affinity for pNPC (Km of 0.680 mM) and was competitively inhibited by cellobiose (Ki of 5.3 mM). A cellulose binding domain was not detected. Putative crystallisation conditions were established with 0.7 M phosphate at pH 6.4.

2.2 Introduction

The examination of lignocellulases is often complicated by the activities of other enzymes produced by the host organism. Although a number of these enzymes have been purified directly from host cultures, it often involves long and complicated purification schemes. One strategy which allows the analysis of individual enzymes is by cloning the encoding genes in heterologous hosts which do not produce any cross reacting enzymes. This allows the properties of recombinant enzymes to be easily examined, even in unpurified samples.

A great number of cellulase and hemicellulase encoding genes have been isolated, with the use of recombinant DNA technology, from a variety of lignocellulolytic organisms and has permitted the examination of these enzymes in both crude and purified extracts. The recombinant enzymes examined include both endo- and exo-glucanases and, more specifically, a few cellodextrinases.

Cellodextrinase encoding genes have been isolated from *Pseudomonas fluorescens* (56) and the two rumen bacteria, *Fibrobacter succinogenes* (76) and *Butyrivibrio fibrisolvens* (26). In addition, the recombinant *F. succinogenes* enzyme has been compared to the native enzyme, purified from both the periplasmic and extracellular fractions, and was shown to be fully functional in *E. coli* and possess similar properties to those of the native enzyme (76,106).

Using recombinant techniques a cellulase gene (*cel*A) was isolated from *R. flavefaciens* by Barros and Thomson (1987) and was thought to encode an endoglucanase by virtue of its activity on carboxymethyl cellulose (CMC). The gene, contained on the plasmid pMEB200, was found to be expressed in the late log phase of growth and the expressed enzyme accumulated in the periplasm of *E. coli* (19). Subsequent studies demonstrated that *cel*A encoded a polypeptide of 39 kDa which was secreted into the periplasm of *E. coli* without a typical leader sequence (219,220). Expression of this gene was under the control of its own

promoter, although no typical -10 and -35 consensus sequences were identified (220).

Preliminary characterisation with crude extracts indicated that CelA may be an exo-acting cellodextrinase (Jennifer Thomson, personal communication), and not an endoglucanase as was first thought. This preliminary characterisation, and the detection of cellodextrinase activity in *R. flavefaciens* cultures (158), prompted a more thorough investigation into the mode of action and biochemical properties of the recombinant enzyme. In the following report the purification and biochemical characterisation of the enzyme encoded by the *cel*A gene will be presented.

2.3 Materials and Methods

2.3.1 Bacterial strains, plasmids and growth conditions.

Escherichia coli strain JM103 [Δ(*lac pro*) *thi*1 *str*A *sup*E *end*A *sbc*B *hsd*R4 (F' *tra*D36 *pro*A,B *lac*I^q *lac*ZΔM15); (134)] was used in these experiments. The strain was grown with vigorous shaking in large flasks to maximise aeration, at 37°C, in 2X yeast tryptone medium (219). *E. coli* JM103 transformed with pMEB200 carrying the *cel*A and ampicillin resistance genes (19) was grown as before, but with ampicillin at a final concentration of 100 μg/ml. *R. flavefaciens* FD-1 (a gift from M.P. Bryant of the University of Illinois, Urbana-Champaign) was grown on non-rumen fluid (NRF) medium (35) at 35°C with ball milled filter paper (0.1% w/v; AECI Ltd. SA) as the major carbon source.

2.3.2 Enzyme preparation and purification.

To obtain the periplasmic fraction, as described by Wang and Thomson (219), NaCl and Tris pH 7.3 were added to a 1ℓ 16 hr culture of *E. coli* JM103 (pMEB200), each to a final concentration of 33 mM. After a 10 min incubation at room temperature the cells were recovered by centrifugation at 5900 *g* at 20°C and resuspended in 10 ml 33 mM Tris pH 7.3 per gram of cells. The cells were osmotically shocked with the addition of an equal volume of TSE (33 mM Tris-Cl pH 7.3, 40 % (w/v) sucrose, 2 mM EDTA) and incubated at room temperature for 5 min. The cells were recovered by centrifugation as before and the periplasmic contents released by resuspension in 20 ml ice cold 1 mM $MgCl_2$ per gram of cells. The suspension was incubated on ice for 10 min and then centrifuged at 10400 *g* for 10 min at 4°C. The supernatant (cell free periplasmic extract) was decanted and stored at 4°C for later purification. Concentrated periplasmic extracts were obtained by filtration using an Amicon PM10 ultrafiltration system and the retentate was stored at -20°C with 0.05% sodium azide.

The purification of CelA from the cell free periplasmic extract was devised

by Thomas Jorgensen (33) and performed at 4°C. The periplasmic fraction (300 ml) was loaded onto a Q-Fast Flow sepharose anion exchange column (Pharmacia; bed volume 300 ml), pre-equilibrated with 33 mM Tris pH 7.3. Proteins were eluted from the column using a linear 0 to 500 mM NaCl gradient in 33 mM Tris pH 7.3 (2ℓ total volume), and collected in 15 ml fractions. Protein was detected at 280 nm using an ISCO model UA-5 optical unit and recorder. Fractions showing activity towards p-nitrophenol-ß-D-cellobiopyranoside (pNPC; section 2.3) were pooled and concentrated on an Amicon PM10 ultrafiltration membrane. The concentrate was dialysed against deionised water, lyophilised and resuspended in deionised water to a final concentration of 20 mg/ml. Concentrated enzyme samples were stored at -20°C.

Crude extracts of *R. flavefaciens* FD-1 cultures were prepared by centrifugation and cell lysis. To obtain the extracellular fraction, cellulose grown cultures of *R. flavefaciens* were centrifuged (10500 *g*) for 20 min at 4°C and the supernatant (extracellular fraction) recovered. The cells were washed in 20 ml phosphate buffered saline (PBS; (182)), harvested by centrifugation as before and resuspended in 10 ml PBS. The cells were ruptured in a French pressure cell three times on ice after which the lysate was centrifuged (12000 *g*) for 10 min at 4°C to remove unbroken cells and cell debris. The lysate (cell free extract) and supernatant fractions were stored at -20°C with 0.05% sodium azide.

2.3.3 Enzyme assays.

All assays were done in duplicate and repeated at least once. Sample and substrate blanks contained distilled water instead of substrate and sample, respectively. Values presented represent averages of the results.

Reducing sugar release was measured by a scaled down version of the dinitrosalicyclic acid (DNS) method as modified by Ghose (1987). One international unit (IU) of enzyme activity was defined as the amount of enzyme liberating one µmole of reducing sugar per minute.

The release of p-nitrophenol from p-nitrophenol derivatives was measured using the pNPC assay described by Deshpande *et al.* (1984), except that samples were analysed spectrophotometrically at 405 nm. Reactions were stopped with the addition of 14% Na_2CO_3. One international unit (IU) of enzyme activity was defined as the amount of enzyme liberating one μmole of p-nitrophenol per minute.

The substrate specificity of CelA was determined using 100 fold concentrated periplasmic extracts isolated from *E. coli* JM103 (pMEB200), as described above. Appropriately diluted samples were incubated with the relevant substrate in PC buffer pH 6.5 (50 mM potassium phosphate, 12.5 mM citric acid) at 41°C and reducing sugar or p-nitrophenyl release was measured using the DNS or pNPC assays. p-Nitrophenyl derivatives were used at a concentration of 3.4 mM and other substrates were used at 1% (w/v); except for oat spelt xylan (2%), phosphoric acid swollen cellulose (2%), carboxymethyl cellulose (CMC, 2%) and cellobiose (0.1%). Substrates were obtained from Sigma (St. Louis, Mo.) except for Avicel (FMC corporation, Philadelphia) and phosphoric acid swollen cellulose which was prepared according to Wood (1988).

The temperature and pH optima of CelA were measured in PC buffer with purified enzyme using the pNPC assay. The pH optimum was measured at 39°C for a pH range over 4 to 10 and the temperature optimum was measured at pH 6.5 for a temperature range over 20 to 55°C. To determine the thermal stability of CelA, samples were held at various temperatures from 4 to 55°C for 30 min and activity was then determined at 41°C. For both the pH and temperature measurements the sample showing the highest activity was taken as 100% and all other samples were expressed as a percentage of the most active sample.

The effects of divalent cations and other substances on CelA activity were determined with purified enzyme by measuring pNPC hydrolysis in 50 mM MES (4-morpholinoethanesulfonic acid) buffer pH 6.5 at 41°C. Samples were centrifuged in a microfuge for 3 min after the reaction to remove any precipitates

that may have formed. The relevant cations were present as chloride salts. Final concentrations were 2.5 and 25 mM, except for $HgCl_2$ (0.5 mM), dithiothreitol (DTT; 1.62 mM), EDTA (5 mM) and glucose (25 mM). Pretreatment of CelA with 10 mM EDTA for 1 hr at 4°C was followed by dialysis against deionised water for 16 hr at 4°C.

The kinetic properties of CelA were determined using the pNPC assay and performed in PC buffer pH 6.5 at 41°C. To determine the initial rates of reaction, 2 ml of enzyme (2 μg/ml final concentration) and 2 ml of substrate, at concentrations from 1.25 to 10 mM, were equilibrated to 41°C for 5 min and then combined. Samples (500 μl) were taken at various time intervals from 0 to 32 min. The initial rates, for the various substrate concentrations, were determined from the linear portion of the reaction. The Km and Vmax were determined from linear regression of the initial rate data, using the method of Lineweaver and Burke (1934). Kinetic experiments were performed both in the presence and absence of cellobiose (25 mM final concentration).

To determine the products of cellodextrin hydrolysis, CelA (0.9 μM) was incubated with 50 μM cellotriose, cellotetraose, cellopentaose or cellohexaose (Merck) in half strength PC buffer (pH 6.5) at 41°C. Samples were taken at various time intervals and the hydrolysis products were separated on a CarboPac PA-100 High Performance Anion Exchange (HPAE) column (Dionex). Sugars were eluted using a 10% to 50% gradient of 0.2 M sodium acetate in 40 mM sodium hydroxide and detected using pulsed amperometry (ED-40 electrochemical detector; Dionex). The products of cellulose hydrolysis were determined similarily, except that CelA (0.9 μM) was incubated with hydrochloric acid treated cellulose (0.25% final concentration; M. Claeyssens, Gent University).

Protein concentrations were determined using the method of Bradford (1976) or the bicinchoninic acid (BCA) reagent (Pierce chemical company; used as directed) with BSA fraction V standards.

2.3.4 Preparation of anti-CelA antibodies

Anti-CelA antiserum was prepared by Barbara von Wechmar (University of Cape Town). To obtain anti-CelA serum, purified enzyme (1 mg) was injected subcutaneously into rabbits with Freund's complete adjuvant and boosted every fourteen days with Freund's incomplete adjuvant. Antiserum was obtained one month after the initial injection. Antibodies were purified by ammonium sulphate precipitation and DEAE-cellulose filtration (40). Anti-*E. coli* antibodies were removed from the immunoglobin preparation by affinity chromatography (182). The method was followed as described, except that *E. coli* cell lysates, in 0.1 M $NaHCO_3$, were obtained using a French pressure cell. The *E. coli* proteins were coupled to cyanogen bromide activated Sepharose 4B (Pharmacia) as described by Colowick and Kaplan (1976).

2.3.5 Polyacrylamide gel electrophoresis (PAGE)

Proteins were separated on 12.5% SDS PAGE gels with 6.4% stacking gels as described by Laemmli (1970). Prior to loading, proteins (1 - 400 µg) were precipitated with 2 volumes of 100% ethanol or acetone and resuspended in 0.25 M Tris pH 6.8. Electrophoresis was performed at room temperature or 4°C at constant current (15 to 40 mA) until the bromophenol blue tracking dye reached the bottom of the gel. The proteins were visualised using the PAGE blue 83 staining method described by Sambrook *et al.* (1989).

2.3.6 Western blotting

Western blotting was performed using the method described by Sambrook *et al.* (1989). Anti-CelA antibodies were used at a 1/250 dilution (4 µg/ml) and in turn were detected with a secondary antibody conjugate (goat anti-rabbit alkaline phosphatase; Sigma) at a 1/5000 dilution. Colorimetric visualisation was performed using the chromogenic substrates, BCIP and NBT (182). Colour was allowed to develop at room temperature and stopped by washing the nitrocellulose membrane in deionised water.

2.4 Results and Discussion

2.4.1 Purification of CelA.

The presence of CelA in the periplasm of *E. coli* was surprising. Despite the lack of a typical leader sequence at its amino terminal end (220), CelA was isolated and purified from the periplasmic fraction of this organism (Table 2.1, Figure 2.1). Although uncommon, the location of recombinant enzymes lacking a signal sequence in the periplasmic fraction of *E. coli* has been noted previously. An example is the recombinant *Butyrivibrio fibrisolvens* cellodextrinase where most of the total activity (55%) was located in the *E. coli* periplasm, despite the lack of a typical signal peptide (26).

Table 2.1 Purification of CelA from *E. coli*

Step	Activity[a] (U/ml)	Protein (mg/ml)	Total activity (kU)	Total protein (mg)	Specific activity (U/mg)	Yield (%)	Purification (fold)
Cell associated	4.3	2.1	4.29	2073	2.1	100	1
Periplasm	2.6	0.5	0.78	142.8	5.5	18.2	2.6
Column filtrate	2.4	0.2	0.36	34.9	10.3	8.4	4.9
PM10 retentate	60.3	5.9	0.33	32.4	10.2	7.7	4.9

[a] Units are μmole pNP released per min

The decreased amount of contaminating protein in periplasmic extracts was desirable and, although only 18.2% of the total cellular CelA activity was isolated from this extract, subsequent purification yielded 32 mg of protein from one liter of culture (Table 2.1). Fractionation of the periplasmic proteins on Q-fast flow sepharose allowed CelA to be eluted as a major sharp peak in the fractionation profile, from 310 to 428 mM NaCl (Figure 2.1). After concentration of the eluted enzyme there was a five-fold increase in the purity, which was verified by SDS PAGE (Figure 2.2).

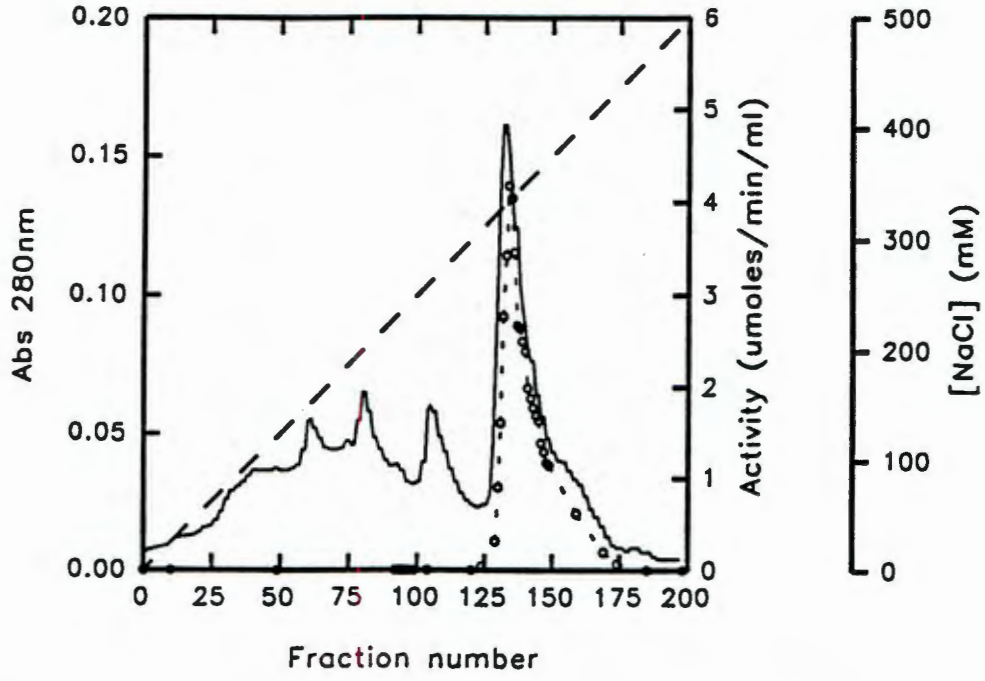


Figure 2.1 Fractionation of periplasmic proteins on a Q-Fast flow sepharose anion exchange column using a 0-500 mM NaCl gradient in 33mM Tris, pH 7.3. Symbols: ——, Abs_{280}; -o-, pNPCase activity; - - - - NaCl gradient.

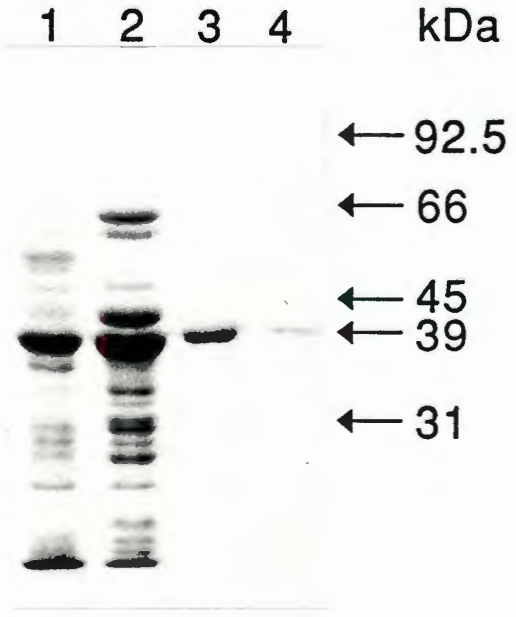


Figure 2.2 SDS-PAGE of the proteins isolated during the purification of CelA. Lane 1, total cell associated protein, lane 2, periplasmic extract; lane 3, Q-Fast flow eluent; lane 4, Amicon PM10 retentate.

Although CelA was readily purified from the periplasm, there was periodic variability in enzyme yields and purity. A number of approaches were tested to alleviate this variability, but all were unsuccessful. One possible explanation arises from the observation that variable expression of CelA occurred in different strains of *E. coli* (219). This, along with the lack of a signal peptide, suggests that the variability in CelA purification may be due to unidentified factors involved in either the expression or transport of the enzyme in the heterologous host.

To confirm the presence of CelA in *R. flavefaciens* FD-1, anti-CelA antibodies were raised, purified (Figure 2.3) and used to probe cell-free extracts (Figure 2.4). A faint band, of similar molecular weight to CelA, was detected in 400 μg of cell-free extract but not in the supernatant fractions (not shown). This result suggested that CelA is a minor cell-associated protein of *R. flavefaciens*.

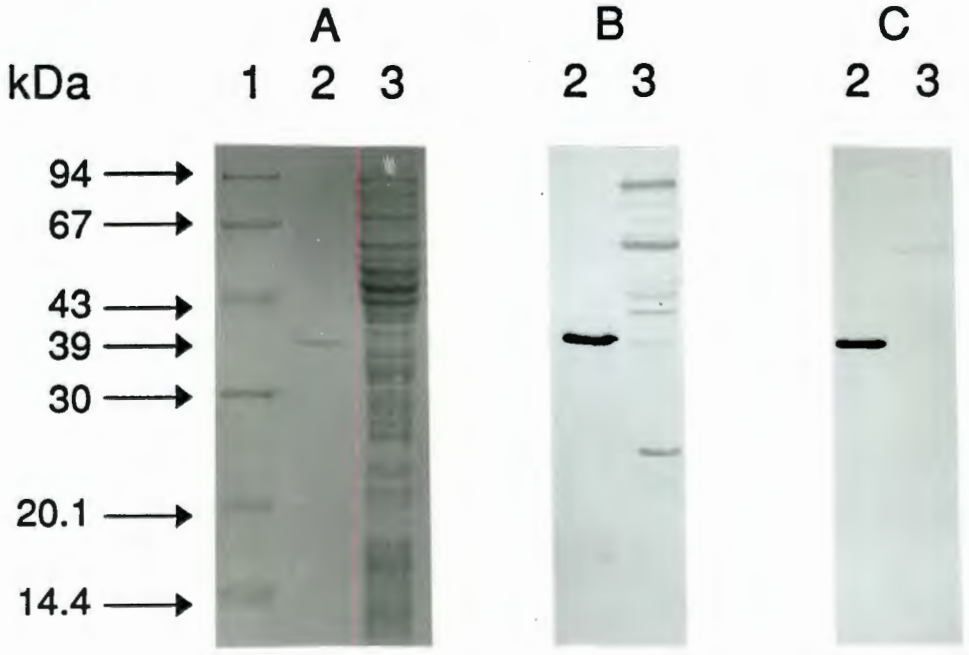


Figure 2.3 Western blots showing purification of anti-CelA antibodies. The SDS-PAGE gel, stained with PAGE blue 83, is shown in (A), ammonium sulphate and DEAE-cellulose purified antibodies were used as probes in (B), while affinity purified antibodies were used as probes in (C). Lane 1, molecular weight markers; lane 2, purified CelA (2 μg); lane 3, JM103 total cell extract (70 μg).

Table 2.2 The activity of CelA on various substrates

Substrate	Specific Activity (IU per mg protein)
pNPC	4.96
Lichenan	0.26
Xylan	0.05
CMC	0.01
Acid swollen cellulose	0.01
Cellobiose	0.01
MUC	+
MUX	ND
pNPX	ND
Laminarin	ND
Avicel	ND
Salicin	ND

+, positive reaction not quantified
ND, not detectable

The activity on cellodextrins, shown by HPAE analysis, demonstrated the exo-type action and dextrinase activity of CelA (Figure 2.5 A to D). The predominant release of cellobiose during the hydrolysis of these dextrins suggests an exo-type mode of action. The production of cellotriose and glucose during the hydrolysis of cellotetraose and cellohexaose was also detected, but the cellotriose was completely hydrolysed by the end of the reaction. The weak endoglucanase and cellobiosidase activities of CelA are probably responsible for the production of cellotriose and glucose during the hydrolysis of these two dextrans.

Further evidence for the exo-mode of action of CelA was obtained by observing the products of cellulose hydrolysis (Figure 2.6). Although CelA does not have high activity on cellulose (Table 2.2) and the reaction proceeded very slowly, the release of cellobiose was observed. The release of cellodextrins other than cellobiose was not observed, further indicating that CelA operates with an exo-type mode of action.

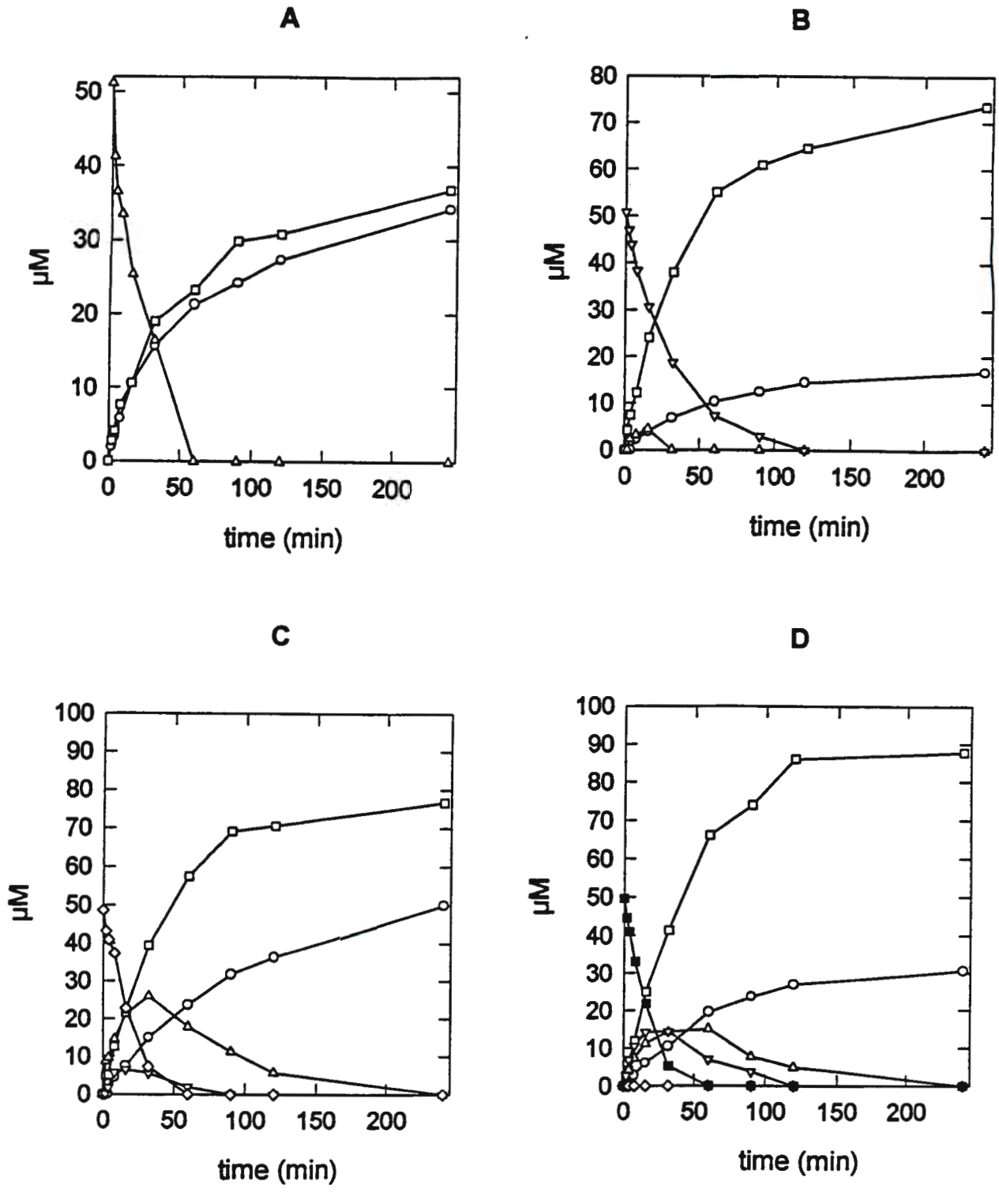


Figure 2.5 HPAE analysis of cellotriose (A), cellotetraose (B), cellopentaose (C) and cellohexaose (D) hydrolysis by CelA. Symbols: ■, cellohexaose; ◇, cellopentaose; ▽, cellotetraose; △, cellotriose; □, cellobiose; o, glucose.

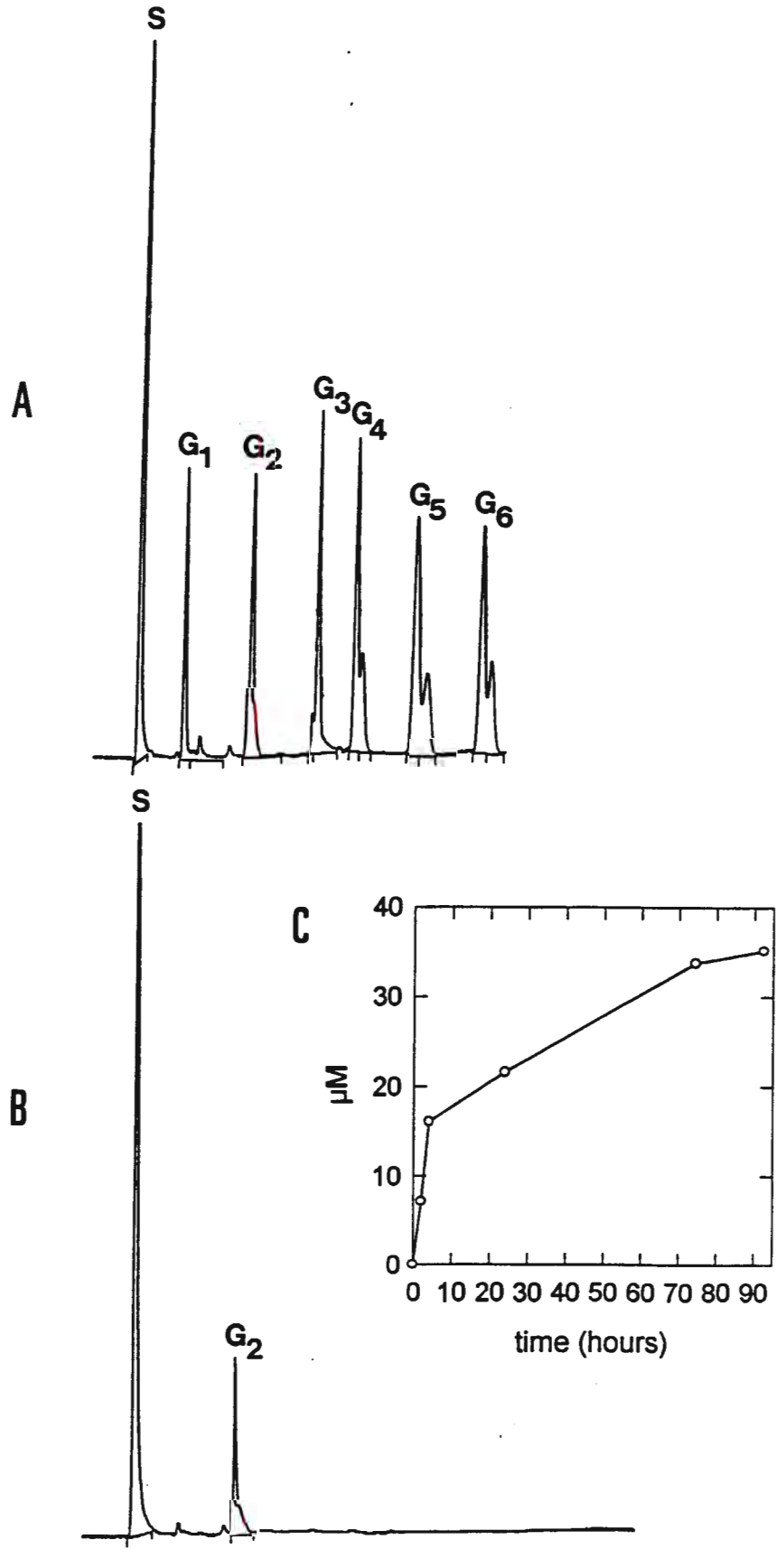


Figure 2.6 HPAE analysis of cellulose hydrolysis by CelA. A, chromatogram of cellodextrin standards (50 μM) from glucose (G1) to cellohexaose (G6); B, chromatogram of cellulose hydrolysis products at 24 hours; C, time course measurement of cellobiose release during the hydrolysis of cellulose by CelA (o, cellobiose). Label S indicates the solvent peak in the chromatograms.

This is not the first example of an exo-acting cellodextrinase from rumen bacteria. The cellodextrinase isolated from *F. succinogenes* attacked cellodextrins with an exo-type action (105), and that from *B. fibrisolvens* also appeared to have exo-type activity (26). In contrast, however, the cellodextrinase from the non-rumen *P. fluorescens* operated in an endo-type fashion (56).

Interestingly, the *B. fibrisolvens* cellodextrinase also displayed multiple activities similar to those of CelA. Endoglucanase and lichenase activity, but barely detectable levels of xylanase activity, were obtained with this enzyme (120). These multiple activities are not uncommon and are probably due to similarities in substrate structure. Other cellulases from *Ruminococcus spp.* with multiple activities have been obtained in clones from *R. flavefaciens* (104) and from *R. albus* (103,154).

2.4.3 Biochemical characteristics of CelA.

The optimal pH and temperature of CelA corresponds to those found in the rumen environment. The pH optimum of CelA was 6.5, with at least 85% activity remaining 0.5 pH units on either side of the optimum (Figure 2.7).

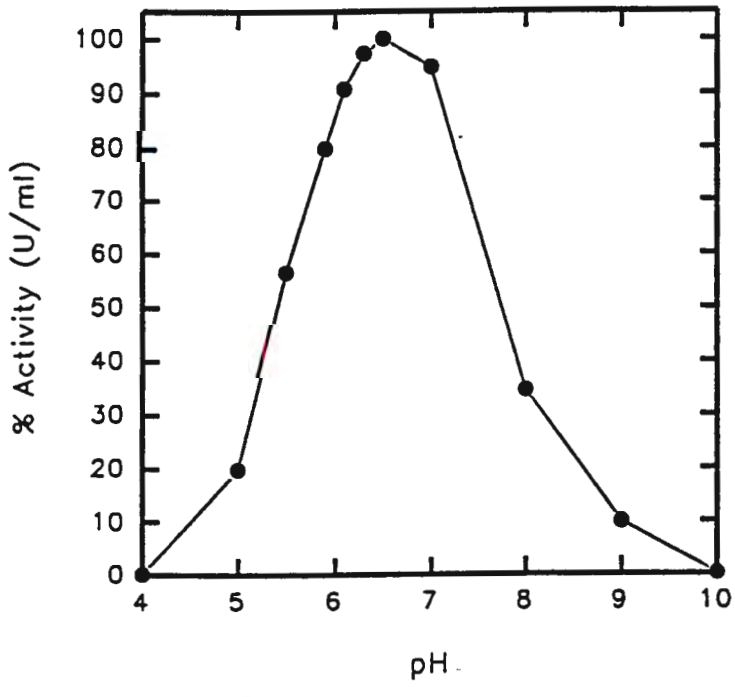


Figure 2.7 pH optimum of CelA measured at 39°C in PC buffer.

Dextrinase activity increased linearly from 20°C to 41°C, where maximal activity was obtained, but only 33% of the maximal activity remained at 50°C (Figure 2.8). The enzyme was stable up to 40°C and by 45°C, 83% of the activity remained (Figure 2.8). No activity was detected by 55°C. There also appears to be a correlation between the thermal optimum of the enzyme and its stability.

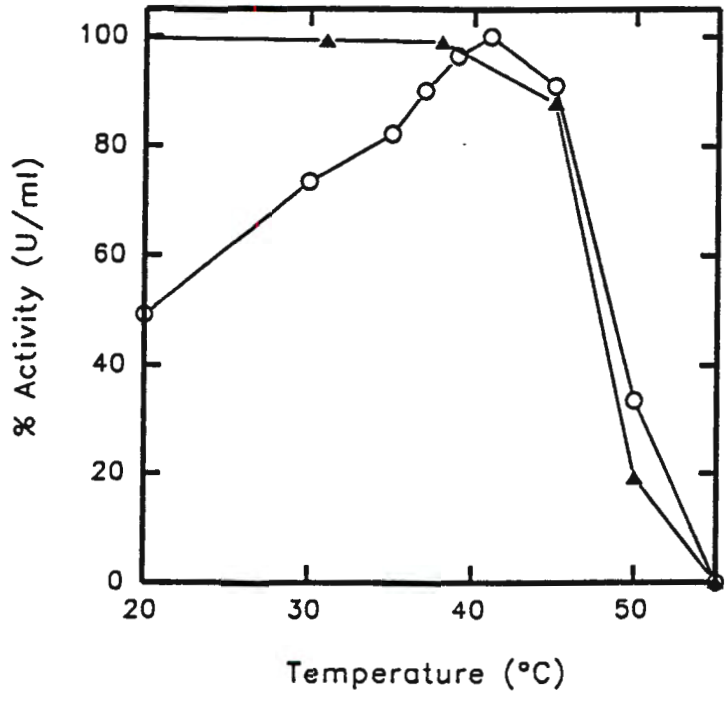


Figure 2.8 Temperature optimum (o) and stability (▲) of CelA in PC buffer pH 6.5.

Pettipher and Latham (1979) described the stimulatory effect of calcium and magnesium ions on cellulases of *R. flavefaciens*, but these two divalent metal cations had no effect on CelA activity (Table 2.3). The transition metals Co^{++}, Ni^{++}, Zn^{++} and Hg^{++} inhibited the enzyme, with Zn^{++} and Hg^{++} completely inactivating the enzyme (Table 2.3). The inhibitory effect of these transition metals, specifically mercury, on enzymatic activity suggests the presence of a sulfhydryl group(s) in

the active site, as demonstrated for an exoglucanase from *Trichoderma reesei* (166). CelA was not affected by EDTA or DTT, indicating that metal ion cofactors or reducing conditions were not required for activity. Pretreatment of the enzyme with EDTA, before addition of the cations, had no significant effect on inhibition levels (results not shown).

Table 2.3 Effect of divalent metal cations and other substances on CelA activity

	% Activity [a]	
	2.5 mM	25 mM
Ca^{++}	93	78
Mg^{++}	90	72
Co^{++}	30	16
Ni^{++}	10	0
Zn^{++}	0	0
Cellobiose	87	45
Glucose	ND[b]	98
Hg^{++} (0.5 mM)	0	
DTT (1.6 mM)	97	
EDTA (5 mM)	96	

[a] Activity expressed as % of control (control activity: 6.2 μmol/min/mg protein).
[b] Not determined.

The kinetic parameters of CelA were determined using pNPC (Figure 2.9). In addition, as the end product cellobiose inhibited CelA activity (Table 2.3) these parameters were determined in both the presence and absence of cellobiose (25 mM final concentration; Figure 2.9). The Km was calculated to be 3.89 mM (S.D. 1.29) and 0.68 mM (S.D. 0.08) and the Vmax to be 2.33 μmoles/min/mg (S.D. 0.53) and 1.89 μmoles/min/mg (S.D. 0.08) with and without cellobiose, respectively. The decreased affinity (Km) for the substrate (pNPC) in the presence of cellobiose, without any significant change in hydrolysing ability (Vmax), suggests that CelA is competitively inhibited by cellobiose (Ki=5.3 mM). Interestingly, an exoglucanase purified from *R. flavefaciens* was also inhibited by cellobiose but had a Km for pNPC (3.08) higher than that of CelA (67), while the *F. succinogenes* cellodextrinase had a Km similar to CelA (0.24 mM) (106).

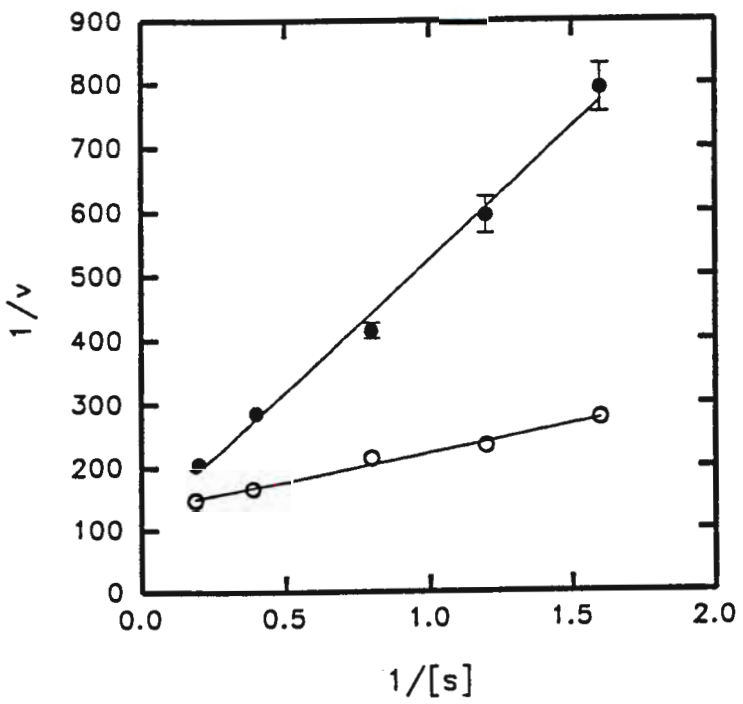


Figure 2.9 Lineweaver-Burke plot showing the kinetic properties of CelA using pNPC as a substrate. The data points are shown with error bars. Symbols: •, CelA incubated with cellobiose; o, CelA incubated without cellobiose.

The inhibition of CelA activity by cellobiose is not unusual and has been demonstrated for other cellulases. For example, an exoglucanase from *Trichoderma*

reesei, and an endoglucanase from *Thermomonospora fusca* were inhibited by cellobiose (166,198). As discussed in Chapter 1, cellobiose can competitively inhibit cellulase activity by binding in the active site.

The amino acid sequence of CelA did not contain regions resembling any of the four typical cellulose binding domains (23), but the enzyme's ability to bind to Avicel was tested in case a novel binding domain was present. A number of cellulases possess cellulose binding domains, including the cellodextrinase cloned from *P. fluorescens* (56), whose function is thought to aid cellulolytic activity on crystalline cellulose (23) (see Chapter 1). CelA, however, did not bind to Avicel suggesting that the enzyme does not possess a cellulose binding domain (results not shown). The lack of such domains was also shown for the cellodextrinases from *F. succinogenes* and *B. fibrisolvens* (26,106).

2.4.4 Crystallisation trials

Although a number of cellulases have been crystallised and their structure determined, no cellodextrinases have been crystallised to date. This prompted an investigation into conditions which would induce CelA to crystallise. Using the hanging drop vapour diffusion technique, needle and flat rhombic crystals were observed using phosphate at room temperature (Figure 2.10). Although the majority were phosphate crystals (Dennis Maeder, personal communication), there were a few potential CelA crystals which differed in size and shape to the phosphate crystals. No crystal development was observed using ammonium sulphate, sodium chloride or MPD at room temperature or 4°C.

To establish the crystallization conditions, a number of parameters of the crystal containing drops were assessed. The drop was tested for pH (6.4), inorganic phosphate concentration (0.7 M) and protein concentration (2 mg/ml). The five fold drop in protein concentration was not indicative of protein crystallisation as the decrease was more likely to have been due to the observed protein precipitates. Although presumptive crystallization conditions were established, the proteinaceous nature of the crystals was not demonstrated and

these experiments were not repeated. As crystallization of CelA was then continued by groups in the UK, the trials were suspended.

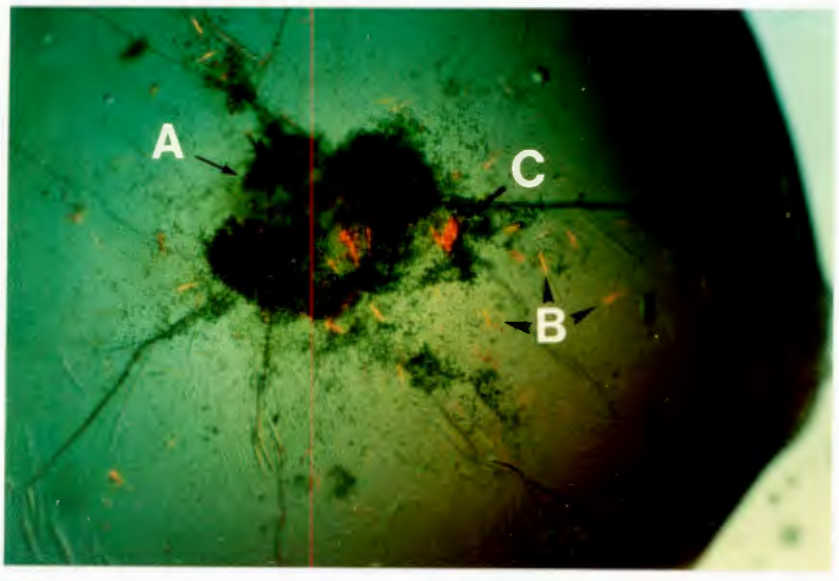


Figure 2.10 Crystallization of CelA using the hanging drop vapour diffusion technique. The drop was photographed under a light microscope with polarising filters. (A) protein precipitates; (B) phosphate crystals; (C) possible CelA crystal.

Thus, CelA is similar to other rumen bacterial cellodextrinases. The enzymes are relatively small and lack cellulose binding domains. Furthermore, they appear to function in an exo-type fashion, have high affinity for dextrin substrates and possess weak multiple activities. From the similarities in function, it would therefore seem that these enzymes play similar roles for their respective host organisms. The function of CelA in *R. flavefaciens* is currently unknown. However, as a cellodextrinase, the enzyme most likely plays the role of a scavenger enzyme, degrading the dextrins released by the activities of exo- and endo-glucanases to cellobiose. The cellobiose which is released from these dextrins would be utilized by *R. flavefaciens* for subsequent growth.

Chapter 3

A search for exocellulase encoding genes in *R. flavefaciens* FD1 and *Clostridium longisporum* B6405

3.2 Introduction

To aid in the identification of exoglucanase activity a number of indicator substrates have been developed. Exoglucanases, or cellobiohydrolases, typically release cellobiose from cellulosic substrates and thus substrates such as pNPC or MUC are normally used where the release of cellobiose produces easily detectable p-nitrophenol or methylumbelliferone (47,214). In some cases, however, endoglucanases and ß-glucosidases have activity on these substrates, which are therefore only used as indicators of exoglucanase activity and are not reliable as classification tools.

Bacterial cellobiohydrolases, or exoglucanases, are not widely found. Only a few have been identified in comparison to the number of fungal exoglucanases (22, 23). Furthermore, exoglucanase activity in the rumen is poorly defined. Endoglucanases, on the other hand, have been identified in every cellulolytic bacterium characterised and many bacteria produce a number of different endoglucanases.

At least two exo-acting cellulases have been identified from *R. flavefaciens* FD-1 including the cellodextrinase, CelA ((33); Chapter 2), and an exo-ß-1,4-glucanase (67). Other potential exoglucanases have been cloned from *R. flavefaciens* strain 186, although true exoglucanase activity was not demonstrated for these clones (104).

Given the presence of exoglucanases in *R. flavefaciens* and the relative scarcity of bacterial exoglucanases in general, a search for novel exoglucanase encoding genes from this organism was prompted. Furthermore, isolation of the gene encoding the exo-ß-1,4-glucanase (67) has not been reported. Although previous attempts to isolate exoglucanase clones from *R. flavefaciens* FD-1 have been unsuccessful in this laboratory (Jennifer Thomson, personal communication), another attempt was made here using a positive selection vector especially constructed for this purpose.

A search for exoglucanase expressing genes was also undertaken in another rumen anaerobe, *C. longisporum*. As discussed in Chapter 1, this organism has recently been reisolated (215) and only one cellulase gene has been characterised to date (137). As *C. longisporum* is capable of growing on a variety of cellulosic substrates (215,217) it must produce a variety of cellulase enzymes. The presence of at least three endoglucanases in this organism (Volker Mittendorf, Ph.D. thesis) further demonstrates the multiplicity of its enzyme system.

The search for exoglucanases in gene banks of both *R. flavefaciens* and *C. longisporum* and the analysis of the clones isolated will be presented in this chapter.

3.3 Materials and Methods

3.3.1 Organisms and growth conditions

E. coli strain K514 (F⁻ *thi*1 *thr*1 *leu*B6 *lac*Y1 *ton*A21 *sup*E44 lambda⁻ rK⁻ mK⁺) and strain K514λ (F⁻ *thi*1 *thr*1 *leu*B6 *lac*Y1 *ton*A21 *sup*E44 lambda⁺ rK⁻ mK⁺) (228) were used for these experiments. *E. coli* strain JM105 (F' *traD36 lacI*^{q} Δ*(lacZ)M15 proA*$^+$*B*$^+$*/thi rpsL* (Str^r) *endA sbcB15 sbcC? hsdR4* ($r_K^-m_K^+$) Δ*(lac-proAB)*) was used for the isolation and maintenance of p41 subclones (see results and discussion) (230). The strains were grown in 2*x* yeast tryptone (YT) broth (182) at 37°C for 16 hr. *E. coli* K514λ was used for maintaining vectors carrying the endonuclease encoding gene and for repression studies requiring the lambda repressor. *R. flavefaciens* was grown as described in section 2.3 except that cellobiose (0.1% w/v; Sigma) was used as the major carbon source.

3.3.2 Standard molecular methods

All standard molecular genetic techniques were followed according to the methods described by Sambrook *et al.* (1989). Restriction enzymes were obtained from Boehringer Mannheim. Digoxigenin (DIG) random primed labeling system (Boehringer Mannheim) was used according to the manufacturer's specifications. The dimethyl sulphoxide (DMSO) method of Chung and Miller (1988) was used to prepare competent *E. coli* cells. Transformed cells were grown with chloramphenicol (20 μg/ml) or ampicillin (100 μg/ml).

3.3.3 Vectors

A positive selection vector, pECORJC1, was constructed from pACYC184 (36) and pECOR251 (a gift from M. Zabeau, Ghent; (234)) (Figure 3.1). The *Acc*I-*Bcl*I fragment of pACYC184, containing the chloramphenicol acetyl transferase (*cat*) gene, was ligated into the *Acc*I-*Bam*HI sites of pECOR251 giving pECORJC0. The ampicillin resistance (*amp*) gene of pECORJC0 was then inactivated by deletion of the *Aat*II-*Pvu*I fragment containing the *amp* promoter and half of the coding region, giving pECORJC1. pECORJC1 contained the endonuclease *ecoRI* gene under the control of the λ rightward promoter, the chloramphenicol

resistance gene and the pBR322 origin of replication. Subcloning experiments were performed using the plasmids pUC18 and pUC19 (230) as vectors.

3.3.4 Genomic DNA isolation

R. flavefaciens genomic DNA, obtained using the method described by Berger *et al.* (1989), was stored in TE buffer at -70°C. *C. longisporum* B6405 genomic DNA was kindly supplied by V. Mittendorf (Department of Microbiology, UCT).

3.3.5 Gene bank construction

R. flavefaciens genomic DNA was partially digested with *Sau*IIIA and fractionated on a 10 to 40% sucrose gradient as described (182). Fractions containing DNA between 3-11 kb were pooled, precipitated and resuspended in TE buffer. These fragments were subsequently ligated into the *Bgl*II site of pECORJC1 and transformed into competent *E. coli* K514 cells. The number of transformants needed to obtain a representative genomic library was calculated from the following formula:

$$N = \frac{\ln (1-P)}{\ln (1-a/b)}$$

where N is the number of clones needed, P is the probability that any given gene is present, *a* is the average insert size and *b* is the size of the *R. flavefaciens* genome (assumed to be similar to the *E. coli* genome, 4 X 10^6 bp). The transformants comprising the library were pooled, the DNA extracted and stored at -20°C.

The *C. longisporum* genomic library (a gift from V. Mittendorf) was constructed using pECOR251 as described (137).

3.3.6 Screening Methods

MUC plates: *E. coli* K514 transformed with pooled genomic library DNA were plated on agar plates containing 0.5 mM MUC (Sigma). After overnight growth at 37°C the plates were visualised under UV (230nm) for fluorescent colonies, indicating cellulase activity.

CMC and xylan plates: *E. coli* K514 transformed with pooled genomic library DNA were replica plated on agar plates containing CMC or xylan (Sigma) at final concentrations of 0.9% (w/v). After overnight growth at 37°C the colonies were removed and the plates stained with Congo red (207). Zones of clearing represented hydrolysis of CMC or xylan and indicated cellulase or xylanase activity.

For liquid assays, *E. coli* K514 transformants were grown in 100 ml 2*x* YT broth as described above. The cells were recovered by centrifugation at 4000 *g* for 7 min, resuspended in 5 ml PBS and ruptured in a French pressure cell on ice. Cell lysates were stored at -20°C with 0.05% NaN_3. Activity against Avicel (1% w/v), acid swollen cellulose (2% w/v) and CMC (2% w/v) were measured using a scaled-up version of the DNS assay described in section 2.3, except that samples were incubated at 37°C for 1 hr. The release of reducing sugar, monitored at 540 nm, was indicative of cellulase activity.

3.4 Results and Discussion

3.4.1 Construction and screening of a *R. flavefaciens* genomic library

A number of unsuccessful attempts to isolate exoglucanase encoding genes from *R. flavefaciens* FD-1 have been made previously (J.A. Thomson, personal communication). Although there are many potential reasons for this, one possibility is lethality due to expression of the heterologous gene in *E. coli*. Exoglucanases are extracellular enzymes in their natural hosts and lethality in *E. coli* could stem from problems in trans-membrane transport (Nigel Minton, personal communication). The trans-membrane transport machinery would be further stressed by the secretion of over-expressed resistance markers, such as ß-lactamase.

To alleviate the potential lethality of overstressing the trans-membrane transport systems, the positive selection vector pECORJC1 was constructed (Figure 3.1). This vector contained the *ecoRI* and the chloramphenicol acetyl transferase (*cat*) genes. As CAT is intracellular the added stress of exported proteins such as ß-lactamase, used as resistance markers, on the cellular export machinery would be alleviated. This vector was then used to create a representative genomic library of *R. flavefaciens* FD-1 by selecting for insertional inactivation of the endonuclease gene. The number of transformants isolated (5521), with an average insert size of 4 kb (results not shown), represented a 99% probability of having any *R. flavefaciens* gene present in the library.

Screening the *R. flavefaciens* genomic library for exoglucanase encoding genes was carried out using MUC, CMC and xylan. The choice of the latter two substrates was prompted by the possibility that exoglucanases may be unable to hydrolyse small substrates, like MUC, but could have weak activities on other longer substrates.

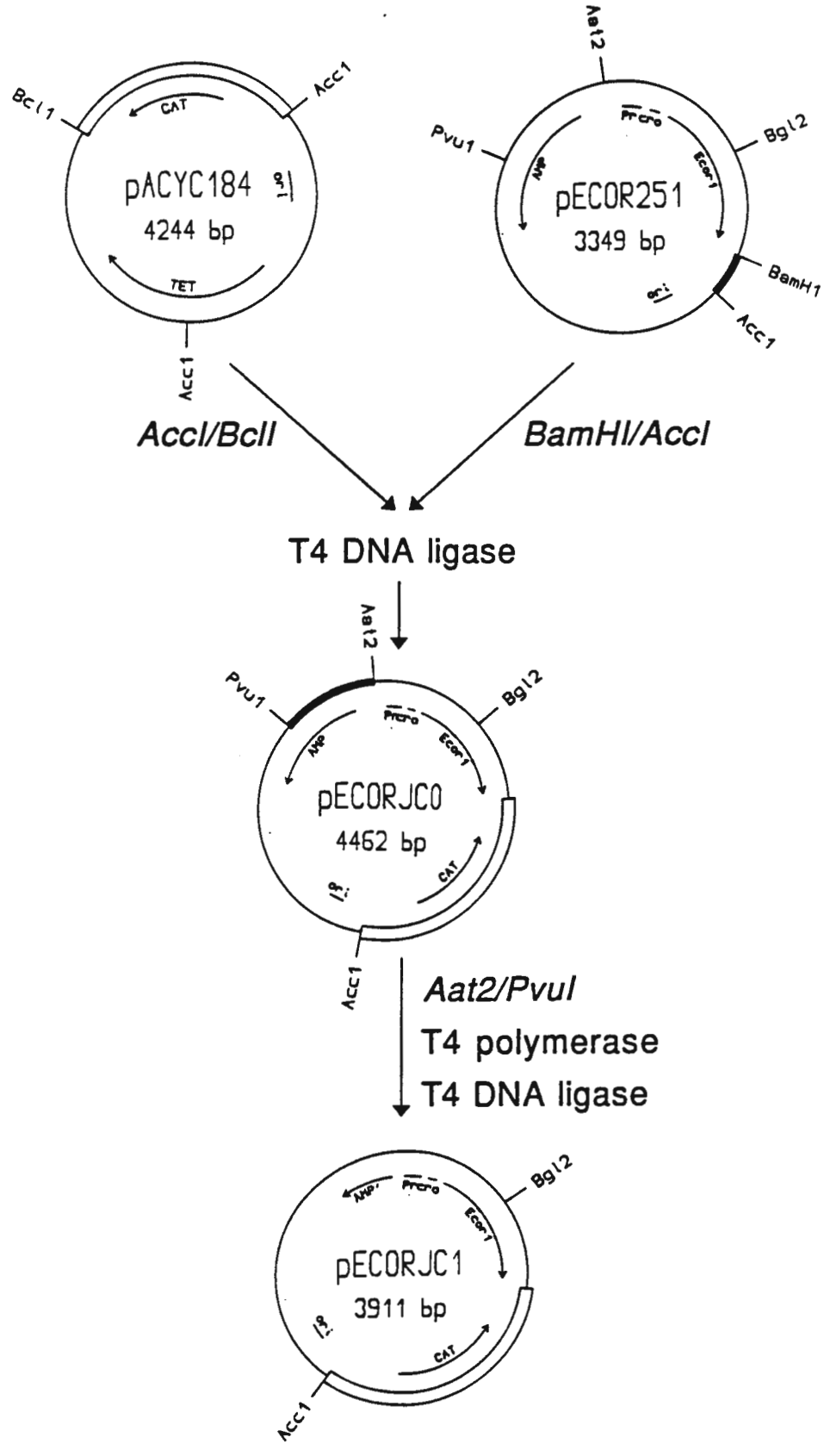


Figure 3.1 Construction of pECORJC1 from pACYC184 and pECOR251. Thick arcs represent portions deleted from the vectors (thin arcs). Only relevant restriction endonuclease sites are shown.

A number of clones were isolated which displayed activity on MUC and /or CMC and five which were active on both these substrates and xylan. All MUC positive clones, however, had restriction banding similarities to the *celA* gene (Figure 3.2). These similarities, both in the restriction banding and substrate activities, suggested that the isolated clones contained the *celA* gene and they were not analysed further. The remaining CMCase positive clones were characterised further with liquid assays using different cellulosic substrates (results not shown). Activity on acid swollen cellulose and possibly Avicel, but low activity on CMC, was taken as suggestive of exoglucanase activity. Although no exoglucanase clones were identified with this approach, at least one putative endocellulase (highly active on CMC) was detected.

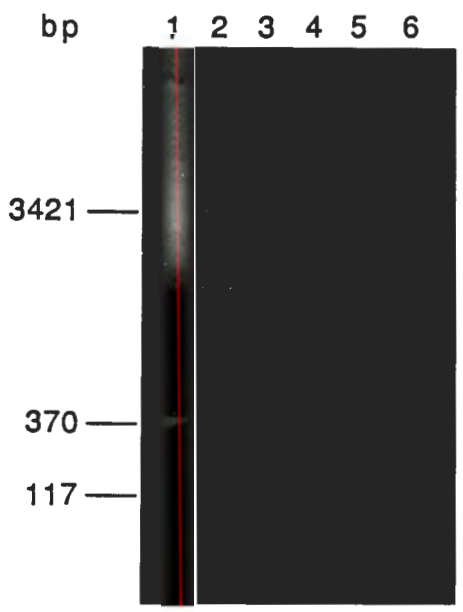


Figure 3.2 Restriction endonuclease banding similarities of MUC positive clones to the *celA* gene. Lane 1, pECORJC1; lane 2, pMEB200; lanes 3 to 6, some MUC positive clones isolated from the *R. flavefaciens* gene bank Plasmid DNA was digested with *HaeII/ Asp700*. The agarose gel was overstained with ethidium bromide to detect the smaller molecular weight bands.

One clone isolated was not active on CMC as was first thought, but produced a yellow pigment in the medium (results not shown). The name *flavefaciens* means yellow producing (100) and it was interesting to note that the production of yellow pigment by *R. flavefaciens* occurs when the organism is grown on cellulose ((100), results not shown). A link between cellulose hydrolysis and pigment production has also been shown for *Erwinia chrysanthemi* (160) and *C. thermocellum* (122).

No further attempts were made to isolate exoglucanase clones from this genomic library. There are a number of possible reasons why an exoglucanase gene was not isolated, including lack of expression of the gene(s) of interest, the inability of the recombinant proteins to form active conformations or associations, a non-representative genomic library, and lethality of the exoglucanase expressing genes in the heterologous host cell. Although lethality may still be due to stress on the trans-membrane transport systems in *E. coli*, the use of internal resistance markers to reduce this potential stress appeared to have no effect. Any future attempts to isolate exocellulase genes from *R. flavefaciens* should take these factors into account.

3.4.2 Screening a *Clostridium longisporum* gene bank

The inability to isolate exoglucanase genes from *R. flavefaciens* prompted a search for exoglucanase genes from another rumen bacterium. As *C. longisporum* is a genetically uncharacterised organism which can utilise a number of cellulosic substrates (215,217), it may produce exocellulase enzymes. Thus a previously constructed genomic library of this organism was screened.

Sixteen positive clones were isolated, using the MUC plate screening method, from approximately 14 000 colonies. *E. coli* K514 which was transformed with plasmid DNA isolated from these clones also displayed MUCase activity, demonstrating that this activity was plasmid encoded (results not shown). Clones 41,42 and 47 had strong activity towards MUC, clones 43, 44, 45, 46 and 48 had moderate activity towards MUC while clones 49, 71, 73, 74, 81, 10, 111 and 112 showed only weak activity towards MUC. Restriction analysis of the plasmid DNA revealed banding similarities between the majority of these clones suggesting that they contained similar genomic inserts (Figure 3.3). Those clones having similar inserts also appeared to have the strongest activity on MUC.

A more detailed analysis of the substrate preferences indicated that none of these clones appeared to have CMCase activity on plates or in liquid assays. This suggested a lack of endoglucanase activity. However, in liquid assays, no activity towards acid swollen cellulose or pNPC was detected while MUCase activity was only detected after prolonged incubation (results not shown). Attempts to establish assay conditions, using higher protein concentrations and different buffer pH's, were unsuccessful. This inability to characterise the clones using liquid assays prompted a genetic analysis.

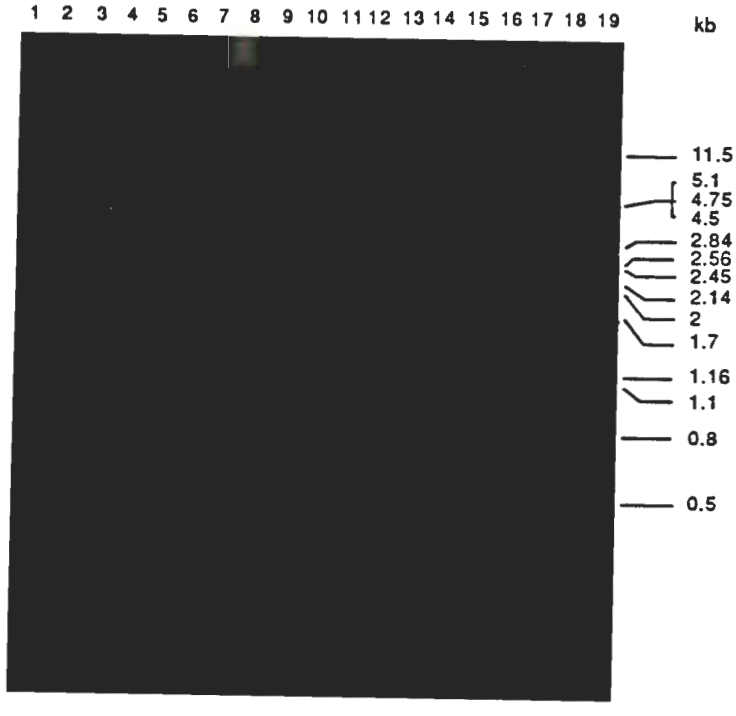


Figure 3.3 Ethidium bromide stained agarose gel showing restriction endonuclease site similarities between the miniprepped plasmid DNA samples of the MUCase positive clones. Plasmid DNA was digested with *HindII/PstI*. Lane 1, λ *PstI* digest; lane 2, pECOR251; lanes 3-11, clones 41-49; lane 12 clone 71; lanes 13-14, clones 73-74; lane 15, clone 81; lane 16, clone 10; lane 17, clone 111; lane 18, pECOR251; lane 19, λ *PstI* digest. Incomplete digestion was obtained with the miniprepped clones.

Plasmid p41, from clone 4.1, was chosen for further characterisation as it shared restriction similarities with the majority of the other clones and encoded strong MUCase activity (Figure 3.3). The plasmid contained an insert of 5.8 kb and was mapped for restriction sites of commonly used enzymes (Figure 3.4). The repression of MUC activity in *E. coli* K514λ, transformed with p41, demonstrated that transcription was running off the vector's λ promotor (Figure 3.5).

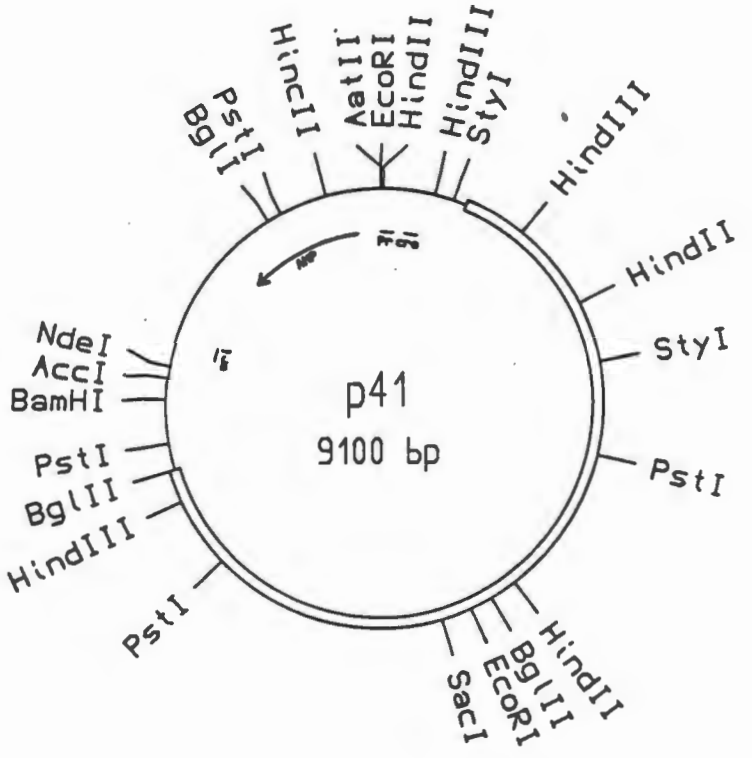


Figure 3.4 Restriction endonuclease map of p41. The open bar represents the *C. longisporum* genomic insert.

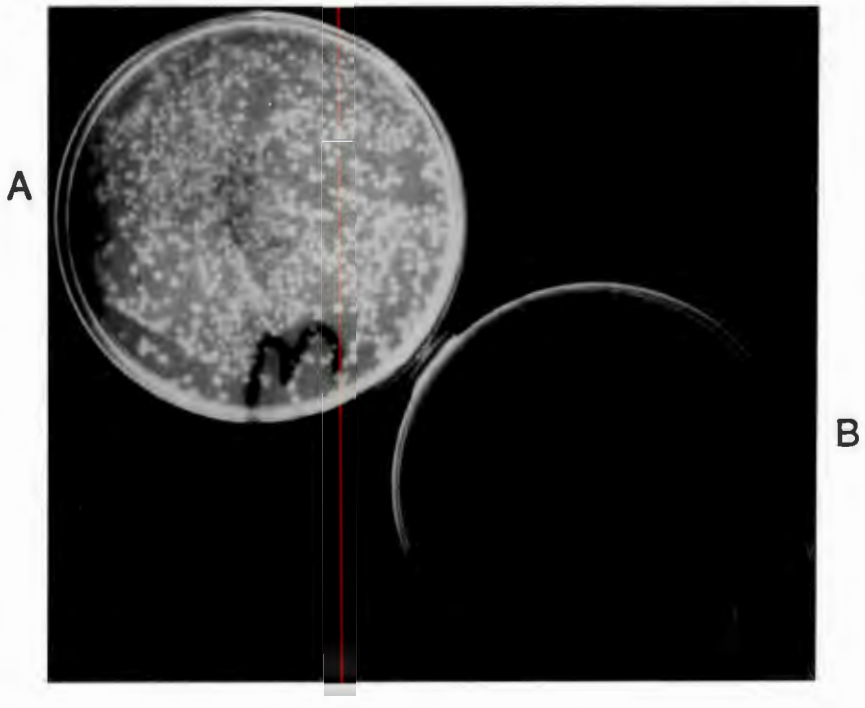


Figure 3.5 Repression of MUCase activity in K514λ. K514 and K514λ were transformed with p41 and plated in YT agar containing 0.5 mM MUC. A, K514 (p41); B, K514λ (p41).

Various subclones of the insert, carried by p41, into pUC18/19 vectors were constructed to isolate the smallest fragment which retained MUC activity (Figure 3.6). As transcription of the insert was running off the λ promoter, the fragments were directionally subcloned downstream of the *lac* promoter of the pUC vectors. Qualitative analysis of the MUCase activities of these subclones showed that pGDB2 was the smallest subclone, carrying a 5 kb insert, which gave full activity. This subclone was constructed from subclones p4.8 and p4.2 by inserting the *PstI* fragment from p4.2 into the *PstI* sites of p4.8. Cells carrying pGDB2 displayed greater activity than those carrying p41 and this was assumed to be due to the higher copy number of the pUC vectors. MUCase activity was lost with any 5' deletions past 500 bp but activity was retained, although appearing to decrease, with all the 3' deletions tested (Figure 3.6).

Southern blotting, using a DIG-labeled 2.5-kb *HindIII* fragment from pGDB2 to probe *C. longisporum* genomic DNA digests, confirmed that the insert originated from the genome of *C. longisporum* (Figure 3.7). Furthermore, the banding patterns observed suggested that a single copy of the insert was present on the chromosome of *C. longisporum*. Southern blot analysis of the original clones having restriction patterns different to p41, with the same *HindIII* probe, revealed homology between all but two of these clones (Figure 3.8). Thus it appears that one particular region of the *C. longisporum* genome was selected for with the screening method used. Also, as MUCase but not CMCase activity was detected, it seemed possible that an exo-acting glucanase had been isolated. As it was not possible to characterise enzyme activity in liquid assays, sequence analysis was chosen to identify the gene(s) conferring the MUC positive phenotype.

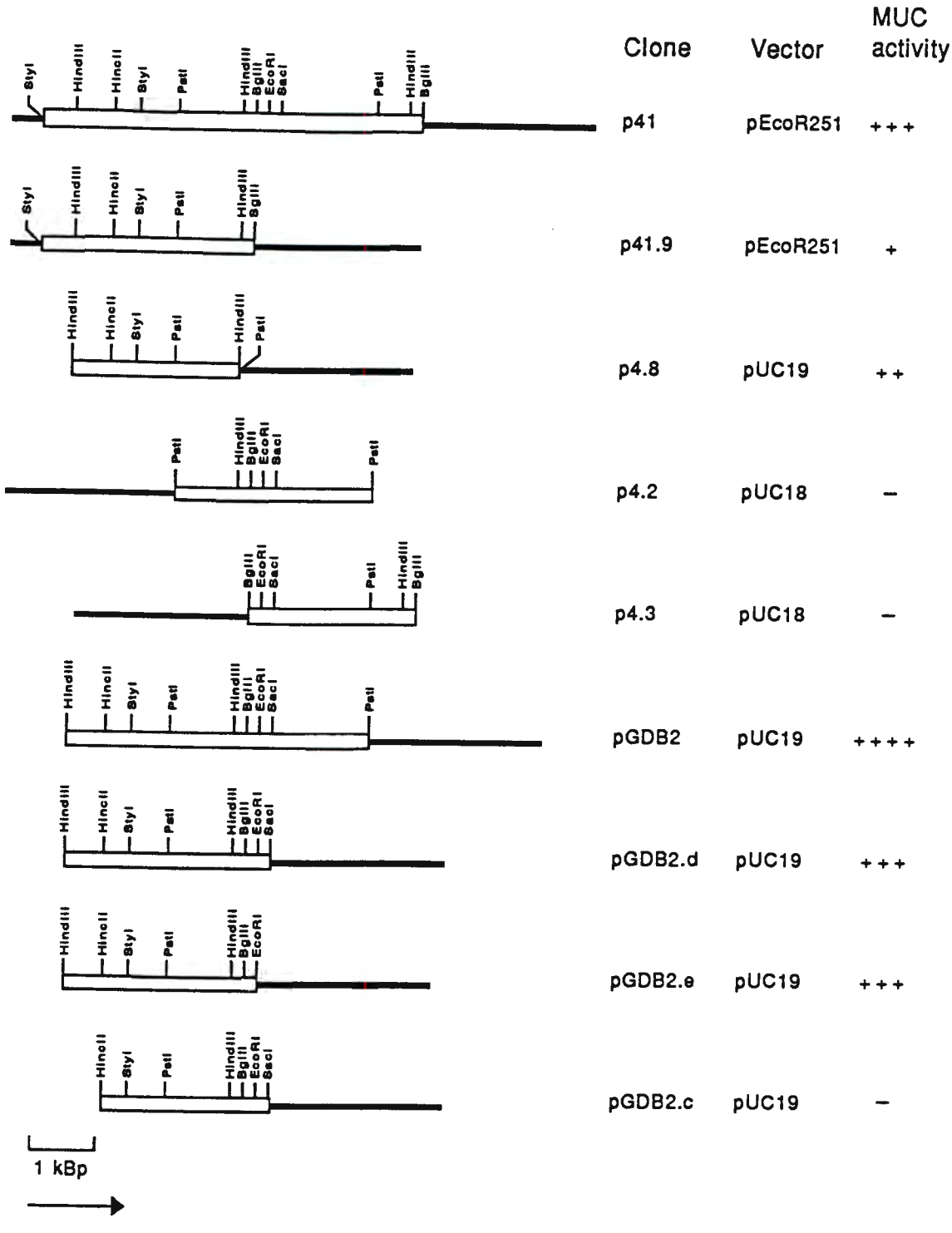


Figure 3.6 Restriction endonuclease map and activity of p41 and subclones. MUC activity was measured qualitatively (- = no activity; + = detectable activity; ++ = weak activity; +++ = moderate activity; ++++ = strong activity). Open bars represent genomic inserts and filled bars represent vector DNA. The arrow indicates the direction of transcription from the vector promoter. Only relevant vector sites are indicated.

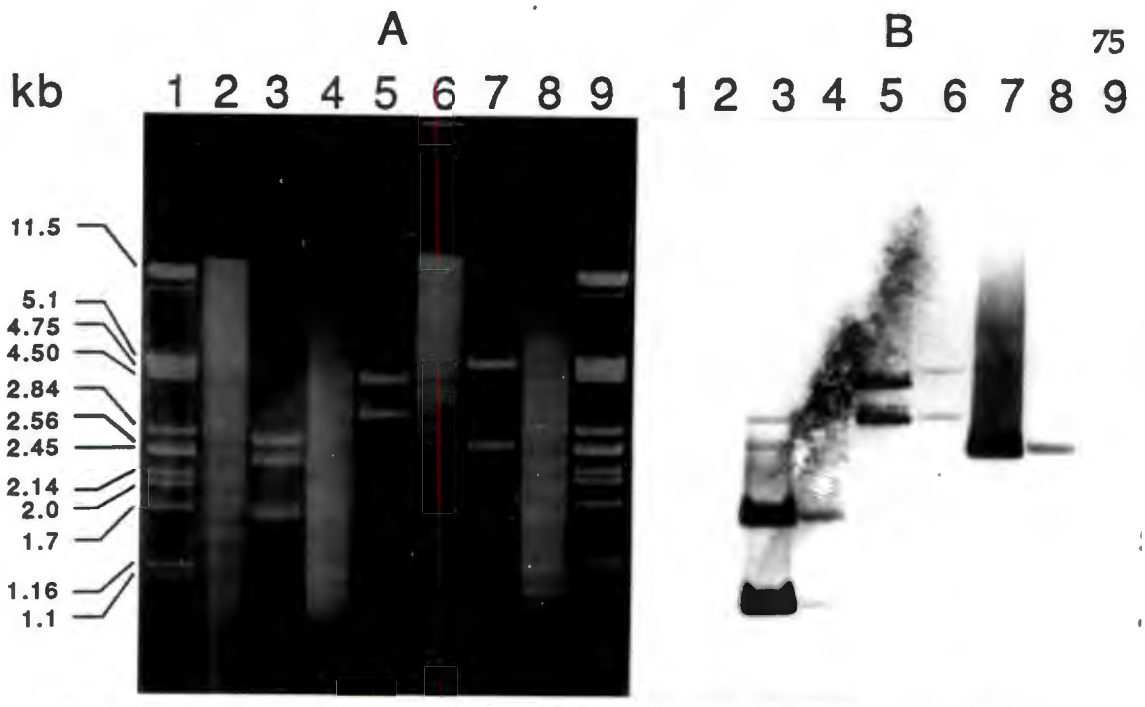


Figure 3.7 Southern blot (B) using a DIG-labeled 2.5-kb *HindIII* fragment from pGDB2 to probe *C. longisporum* chromosomal DNA. The ethidium bromide stained agarose gel is also shown (A). Lane 1, λ *PstI* digest; lane 2, *E. coli* chromosomal *PstI/HindIII* digest; lane 3, pGDB2 *PstI/HindIII* digest; lane 4, *C. longisporum* chromosomal *PstI/HindIII* digest; lane 5, pGDB2 *PstI* digest; lane 6, *C. longisporum* chromosomal *PstI* digest; lane 7, pGDB2 *HindIII* digest; lane 8, *C. longisporum HindIII* digest; lane 9 λ *PstI* digest.

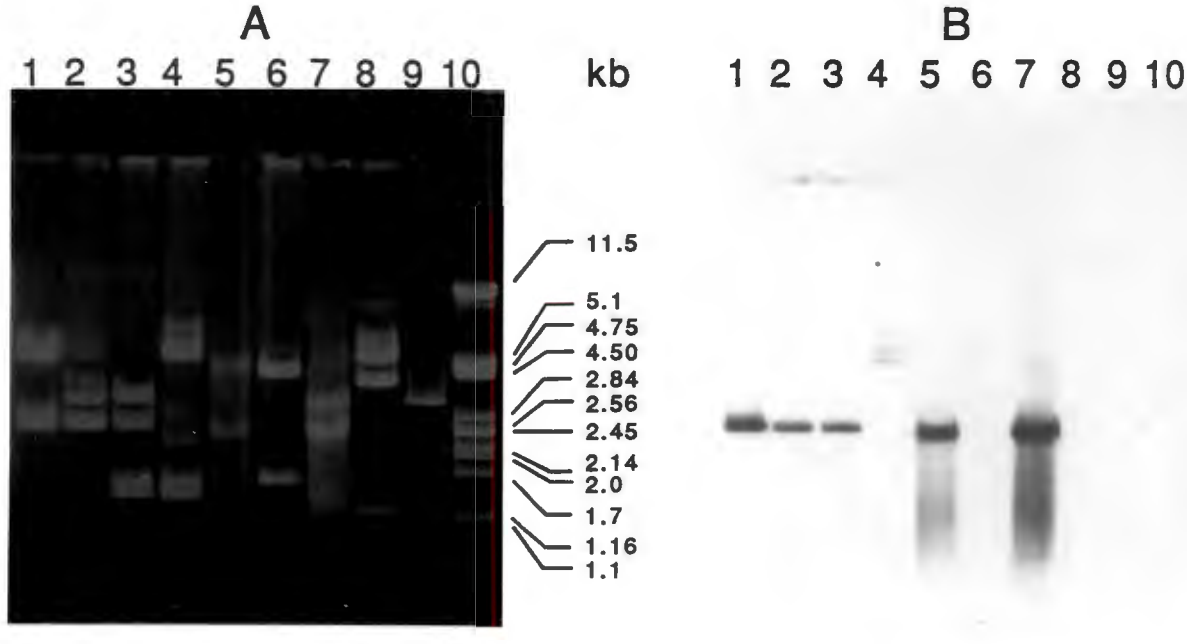


Figure 3.8 Southern blot (B) using a DIG-labeled 2.5-kb *HindIII* fragment from pGDB2 to probe plasmid DNA from other MUCase positive clones which were isolated. The ethidium bromide stained agarose gel is also shown (A). All DNA, unless otherwise stated, was digested with *HindIII*, prior to electrophoresis and subsequent blotting. Lane 1, pGDB2; lane 2, clone 41; lane 3, clone 81; lane 4, clone 71; lane 5, clone 10; lane 6, clone 74; lane 7, clone 111; lane 8, clone 73; lane 9, pECOR251; lane 10, λ *PstI* digest.

Chapter 4

Sequence analysis of the MUC positive clone, pGDB2: isolation of genes involved in aromatic-ß-glucoside uptake and utilisation

4.1 Abstract

Sequence analysis was undertaken to identify the genes responsible for the MUCase activity on pGDB2 (Chapter 3). Four complete and one truncated open reading frames (ORF) were detected and appeared to encode one truncated and two complete proteins for aromatic-ß-glucoside uptake and utilisation. Two downstream genes, encoding a protein of unknown function (ORF6) and a P_{II} nitrogen regulatory protein (*glnB*), were also detected. Our interest in carbohydrate uptake and utilisation in the rumen prompted further investigation into the gene system encoding aromatic-ß-glucoside uptake. The upstream regulatory regions, including the missing 5' portion of the truncated ORF, were then isolated and sequenced. Analysis indicated that regulatory regions, the complete 5' gene of the carbohydrate utilisation system, as well as a truncated gene encoding a methyl-accepting chemotaxis protein (*ΔmacA*) and a T-Box tRNA synthetase gene (*trsA*), had been isolated. The aromatic-ß-glucoside uptake and utilisation genetic system, possibly constituting an operon, includes an antiterminator (*abgG*), a PTS-dependent transmembrane permease (*abgF*) which confers the MUCase activity to *E. coli*, and a phospho-ß-glucosidase (*abgA*). This is the first such system isolated from a Gram positive organism and it shows remarkable similarity, both in sequence and in organisation, to previously characterised *Enterobacteriaceae* genes. As *Clostridium longisporum* is genetically uncharacterised, a comparative analysis with the other isolated genes and regulatory regions was also undertaken.

4.2 Introduction

The lignocellulolytic abilities of *C. longisporum* are poorly characterised, and although the organism is capable of growing on a variety of lignocellulosic materials only one cellulase, an endoglucanase, has been characterised to date (137,215,217). Our initial interest in rumen exocellulases prompted a search for exocellulase expressing clones in a genomic library of this organism (Chapter 3). One clone, carrying p41, was isolated which possessed MUCase activity, but not CMCase activity (suggesting possible exocellulase activity). The inability to characterise the enzyme(s) responsible for this activity in liquid assays (Chapter 3), prompted a genetic analysis. Sequence analysis of cellulase enzymes has identified many components of the enzymes themselves, such as their modular structures, as well as allowing their classification. With the huge repositories of sequence data now available it was hoped that a comparative analysis would allow an identification of the gene(s) present on the genomic insert responsible for the MUCase activity.

Carbohydrate uptake and utilisation in the rumen has been of interest to researchers for a number of years (see Chapter 1). As one of the ultimate aims is to improve fibre digestion in the rumen, and hence animal productivity, an understanding of the systems involved in the metabolism of lignocellulosic substrates is important. Furthermore as lignin is not degraded in the rumen (37) any system involved in the utilisation of phenolic compounds, such as phenolic glycosides, is also of interest. In this chapter the genes on pGDB2, responsible for the MUCase activity, will be shown to be part of an aromatic-ß-glucoside uptake and utilisation system. This exciting discovery prompted further investigation of these genes as well as their regulatory regions. Furthermore, as *C. longisporum* is genetically uncharacterised, an identification and comparative analysis of the genes isolated which flank the aromatic-ß-glucoside gene system, was undertaken.

4.3 Materials and Methods

4.3.1 Bacterial strains and growth conditions

E. coli strain JM105 (F' *traD36 lacI*^{q} Δ(*lacZ*)*M15 proA*$^+$*B*$^+$/*thi rpsL* (Str^r) *endA sbcB15 sbcC? hsdR4* ($r_K^-m_K^+$) Δ(*lac-proAB*)) and *E. coli* strain JM109 (F$^-$ *traD36 lacI*^{q} Δ(*lacZ*)*M15 proA*$^+$*B*$^+$/*e14*$^-$ (*McrA*$^+$) (Δ*lac-proAB*) *thi gyrA96* (*Nal*^{r}) *endA1 hsdR17* (r_k^- m_k^+) *relA1 supE44 recA1*) were used for these experiments. The strains were grown in 2X YT (182) at 30°C or 37°C.

4.3.2 Plasmids and molecular methods

Plasmids used in these experiments are listed in Table 4.1. Plasmid DNA was purified using the Nucleobond purification system (Machery-Nagel GMBH and Co.), as described by the manufacturers and stored in TE buffer at 4°C or -20°C.

Table 4.1 Plasmids used in these experiments.

Plasmid	Markers and Genes[a]	Reference
pGDB2	Amp^r (Δ*abgG*) *abgF abgA* ORF6 *glnB*	Chapter 3
p4.2	Amp^r (Δ*abgF*) *abgA* ORF6 *glnB*	Chapter 3
p4.9	Amp^r (Δ*abgG*) *abgF* (Δ*abgA*)	This Chapter
p81	Amp^r *macA trsS abgG abgF* (Δ*abgA*)	This Chapter
p8B1	Amp^r (Δ*macA*) *trsS abgG* (Δ*abgF*)	This Chapter
pUC18	Amp^r *lacZ*	(230)

[a]Determined from sequence data

Competent bacterial cells were prepared according to the DMSO method of Chung and Miller (1988). Transformed cells were grown as described with the addition of 100 μg /ml ampicillin (Sigma).

All standard molecular techniques were performed according to Sambrook *et al.* (1989). Restriction enzymes were obtained from Boehringer Mannheim and

used according to the manufacturers specifications. Nucleotide sequencing was performed with the Sequenase kit (version 2.0; U.S. Biochemicals), using a modification of the dideoxy chain termination method of Sanger *et al.* (1977). Primers used for sequencing are shown in Figure 4.1.

The method of Henikoff (1984), employing ExoIII, was used to generate sets of nested deletions for sequencing. To sequence pGDB2, forward sets of nested deletions were generated from p4.9 and p4.2 while reverse sets were generated from pGDB2 (Table 4.1). Shortenings were generated from vector *KpnI* and *XbaI* sites, for both the forward and reverse sets (Figure 4.2). To sequence p8B1, shortenings were generated from *SacI/ BamH1* for the forward direction, and *PstI/ StyI* for the reverse direction (Figure 4.4).

```
5'-CAGCACTGACCCTTTTG-3'          -40 primer
5'-AGCGGATAACAATTTCACACAGG-3'  1233 primer
5'-TTTAGTAGAACCAAAGGC-3'         P1
5'-TAGAAGAAATGAAAGCTC-3'         P2
5'-GAAGTAAGAGAGACTATTG-3'        P3
```

Figure 4.1 Primers used for sequencing. The -40 and 1233 primers are pUC universal primers. P1, P2 and P3 primers were synthesised by the Biochemistry Department (UCT) and used for sequencing part of p8B1.

4.3.3 Computer analysis

Computer analysis was performed using the Genetics Computer Group (68) and the Genepro sequence analysis suite of programs (162). Homology searches were performed using BLAST and accessing all protein databases at the National Centre for Biotechnology Information (NCBI) (141). RNA folding and minimum free energies were determined using the FOLD program of GCG, based on the method of Zucker as modified by Freier (62,237,238). Protein hydropathy was determined using the PEPPLOT program of GCG, based on the method of Kyte and Doolittle (1982).

4.4 Results and Discussion

4.4.1 Sequence and analysis of pGDB2

The lack of usable restriction sites necessitated the use of three plasmids for the generation of both forward and reverse sets of nested deletions for sequencing (Figure 4.2). p4.2, initially used to generate pGDB2 (Chapter 3), and p4.9 which is similar to p4.8 (Chapter 3) but with the *HindIII* fragment in the reverse orientation, were used to generate deletions of the "forward" strand while pGDB2 was used for the "reverse" strand.

Sequencing of both strands revealed that pGDB2 contained a genomic insert of 5008 bp (Figure 4.3, Figure 4.6). Four complete open reading frames were detected and a 5' truncated ORF. Transcription of each of these ORFs appeared to run 5' to 3' which was consistent with the observation that transcription was running off the vector promoter (Chapter 3). The coding regions for all the ORFs utilised presumptive AUG start codons, and were closely preceded by potential ribosome binding sequences (193). Due to the high percentage of AT nucleotides, -10 or -35 regions could not be predicted. In addition, three stemloop structures were detected, which may act as putative *rho*-independent terminators (Figure 4.6) (165).

Searches for homology with predicted protein sequences of each ORF indicated that genes encoding phosphoenolpyruvate dependent aryl-ß-glucoside (*abg*) uptake and utilisation had been isolated (Table 4.2). Based on homology analysis, these genes appeared to encode a truncated antiterminator (*abgG*), a transmembrane transport protein (*abgF*) and a phospho-ß-glucosidase (*abgA*). These genes were followed by a stem-loop terminator like structure, mentioned previously (Figure 4.6). Of the two downstream genes only the 3' gene had homology to sequences in the database, appearing to code for a nitrogen regulatory protein (*glnB*), which was also followed by a stem-loop type terminator.

The involvement of the *abgF* gene product, the transmembrane enzyme II, in the observed MUCase activity is apparent when considering the activity of the subclones isolated from p41 and pGDB2 (Chapter 3, Figure 3.6). MUCase activity was retained in all clones which had a complete *abgF* gene despite having deletions of the *abgA* gene, encoding the phospho-ß-glucosidase, and downstream genes. This is discussed further in Chapter 5.

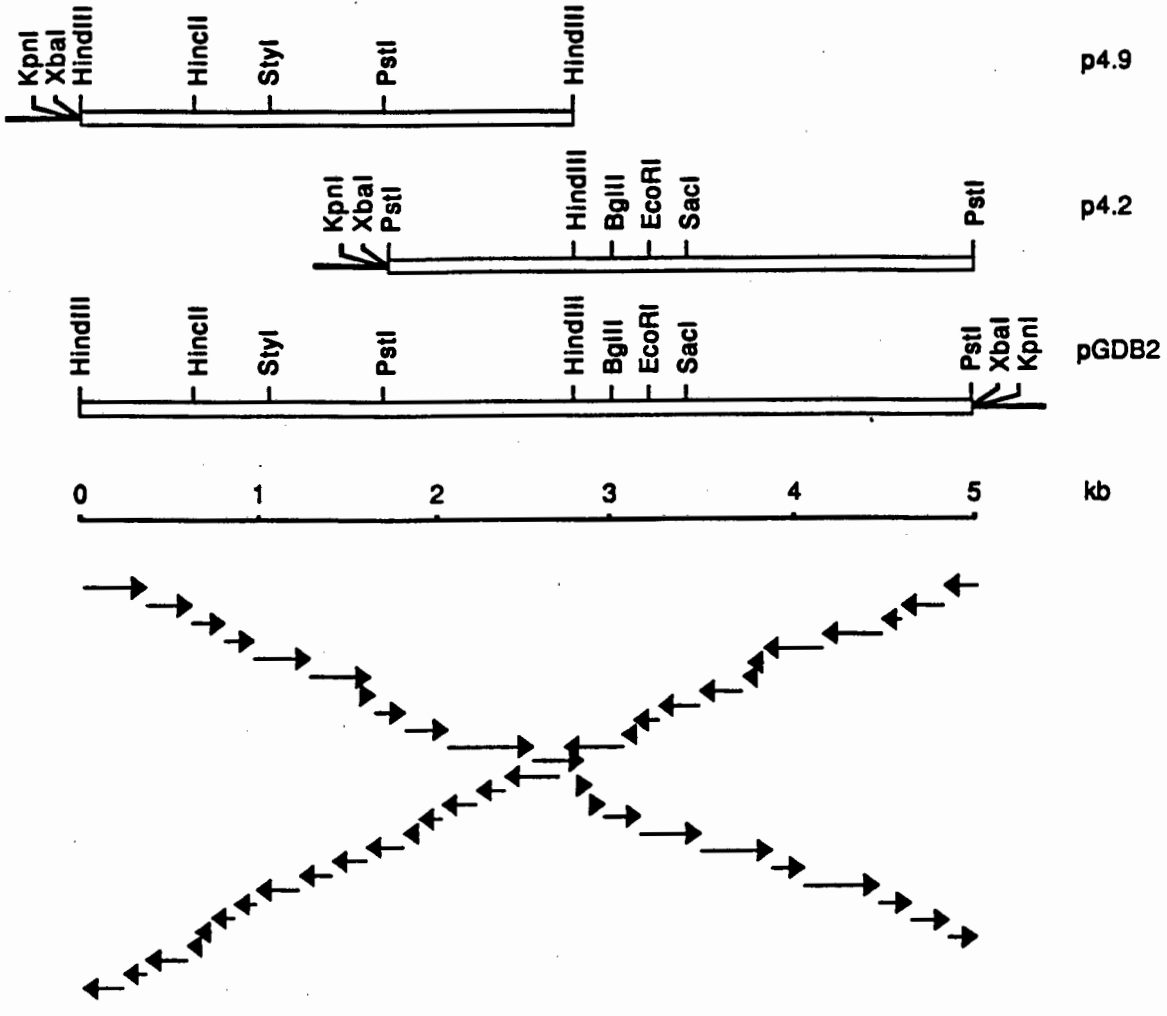


Figure 4.2 Sequencing strategy using pUC universal primers. Nested deletions were achieved using the Henikoff shortening method (90) as described. Forward shorts were obtained using p4.9 and p4.2, while reverse shorts were obtained using pGDB2. Arrows indicate direction and length of sequence contributed by each deletion. Only relevant vector restriction sites are shown.

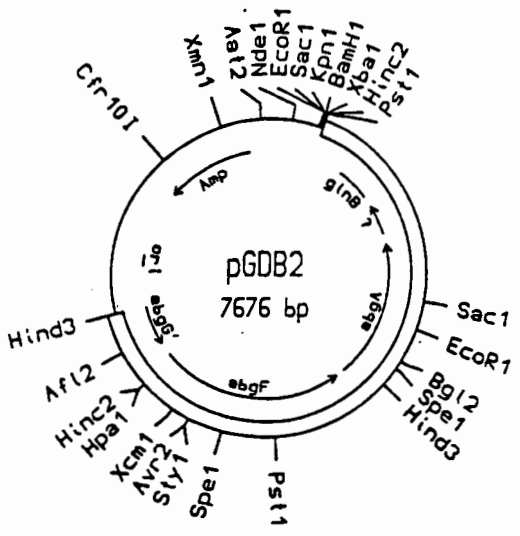


Figure 4.3 Plasmid map of pGDB2 showing restriction endonuclease sites and gene positions. The vector promoter, located 5' to *abgG*, is not shown. The open arc represents the *C. longisporum* genomic insert, and the single line represents vector DNA.

The isolation of putative phosphoenolpyruvate dependent transferase genes using a cellulase indicator substrate was surprising. During the course of these experiments, however, Lai and Ingram (1993) isolated a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* using MUC, when searching for cellulase activity in a gene bank of this organism. It therefore appears that this substrate can also be utilised as a marker for cloning cellulosic PTS uptake systems.

The unexpected nature of this discovery and the possibility of gaining further insight into the mechanisms of carbohydrate utilisation in the rumen, especially those involved in the utilisation of phenolic glycosides, prompted further characterisation of these genes. Furthermore, these would be the first aromatic-ß-glucoside uptake and utilisation genes isolated from a Gram positive organism. Only three such systems have been isolated previously, all from the Gram negative *Enterobacteriaecae* (81,85,192).

4.4.2 Isolation and sequencing of upstream regions

To obtain the complete operon, it was necessary to isolate the upstream region, containing the remainder of the 5' truncated gene, which had shown

homology to previously identified antiterminators, and potential regulatory regions. As a number of the original MUCase positive clones appeared to contain similar regions of the *C. longisporum* genome (Chapter 3), they were analysed further to determine whether any contained the upstream regions of interest.

Based on the initial data from the Southern blot and restriction analysis (Figures 3.3 and 3.8), clone p81 appeared to have the desired upstream regions. It was partially mapped and was found to contain an insert of 8.5 kb (Figure 4.4). Its 3' end (2.7 kb) had a similar restriction pattern to the 5' end of pGDB2, from the *HindIII* to the *BglII* site. The hybridisation of the 5' *HindIII* fragment of pGDB2 to p81, demonstrated by Southern blot analysis, confirmed the similarity of this region (Figure 3.8).

The *XbaI/ PstI* (4 kb) fragment of p81, potentially containing the 5' regions of the operon, was cloned into pUC18 behind the *lac* promoter, creating p8B1 (Figure 4.4). This construct, containing a truncated *abgF* gene, did not possess MUCase activity (results not shown), which is further evidence for the role of enzyme II (AbgF) in MUCase activity (see Chapter 5). Nested deletions of p8B1, in both directions, were generated (not shown) and sequenced as described. Sequencing confirmed the overlap between p8B1 and pGDB2. The full DNA sequence (7.5 kb) of the aryl-ß-glucoside genes, as well as the upstream and downstream regions, derived from p8B1 and pGDB2, is shown in Figure 4.6.

Problems were, however, encountered in the isolation and maintenance of strains carrying p81 and p8B1. Rearrangements of these plasmids was observed to occur in JM105 in large scale liquid cultures which was only alleviated when using the *recA*$^-$ JM109 strain (results not shown). The reasons for the rearrangements were not determined, but possibly arose due to recombination with homologous regions on the *E. coli* chromosome (see below). In addition, growth of bacteria transformed with either of these two plasmids was poor and viability was slightly increased when the cells were grown at 30°C (results not shown).

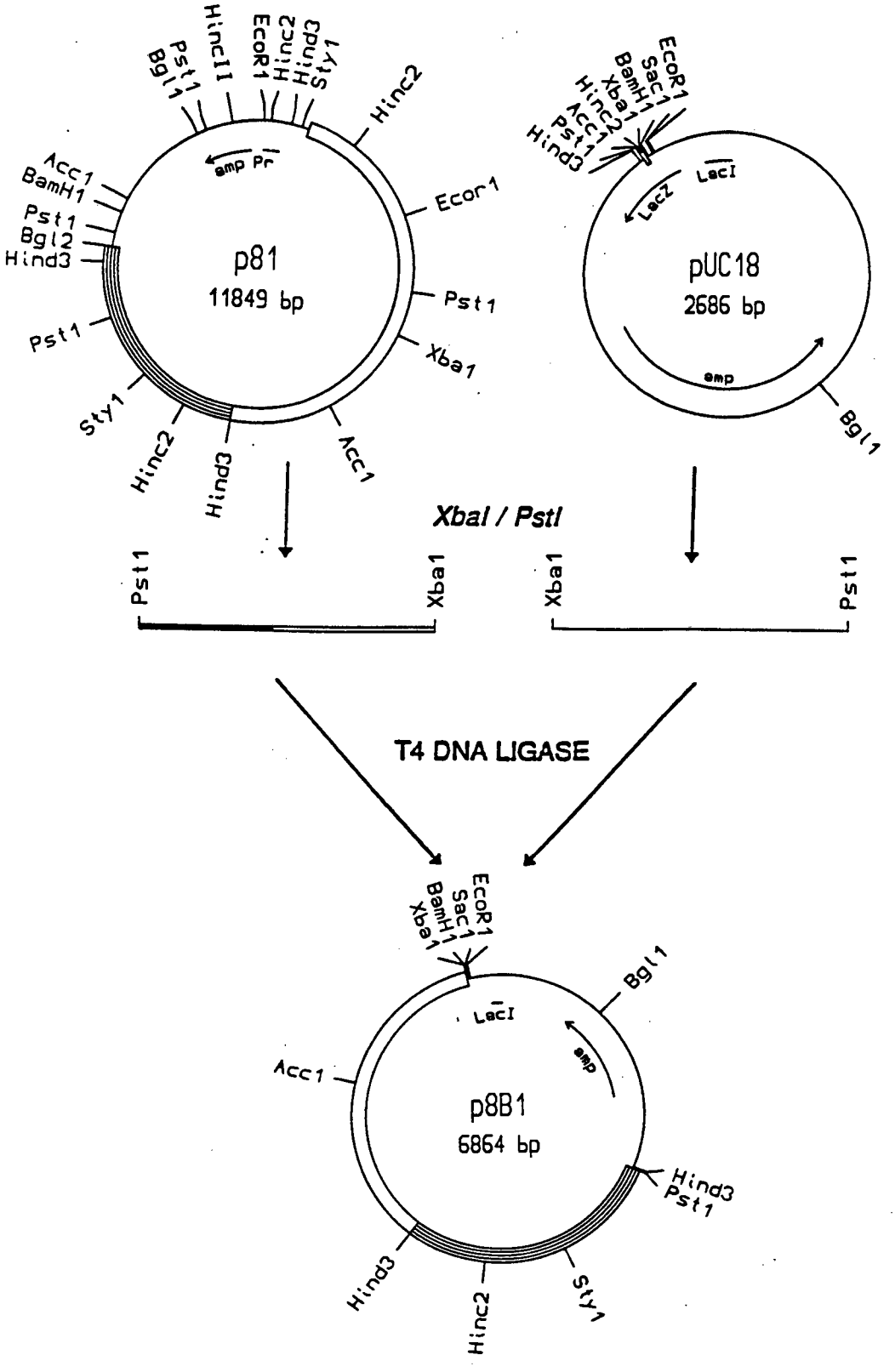


Figure 4.4 Plasmid map of p81 and construction of p8B1. Thick arcs represent *C. longisporum* genomic DNA while thin arcs represent vector DNA. The shaded area indicates the region of overlap with pGDB2. Not all the *Hinc2* and *Hind3* sites are shown.

Two ORFs and two truncated ORFs were located on the genomic insert of p8B1 (Figure 4.5). Based on homology of the predicted amino acid sequence of each of these ORFs it appeared that, reading in the 5' to 3' direction, the 3' end of a methyl-accepting chemotaxis receptor protein gene (*macA*), a tryptophanyl-tRNA synthetase gene preceded by a 5' ORF (*trsA*), the antiterminator gene (*abgG*) and a 5' portion of the membrane transport protein gene (*abgF*) had been isolated (Table 4.2).

Translation of the tRNA synthetase (*trsA*), but not the antiterminator (*abgG*), appears to be initiated from an AUG start codon. Based on homology (see below) translation of *abgG* appears to be initiated from an AUU (ILE) codon which, although rare, has been reported previously (171). The sequence surrounding this codon was rechecked three times with independent deletions in the forward direction and twice in the reverse (results not shown). Both genes were preceded by potential ribosome binding sequences (Figure 4.6).

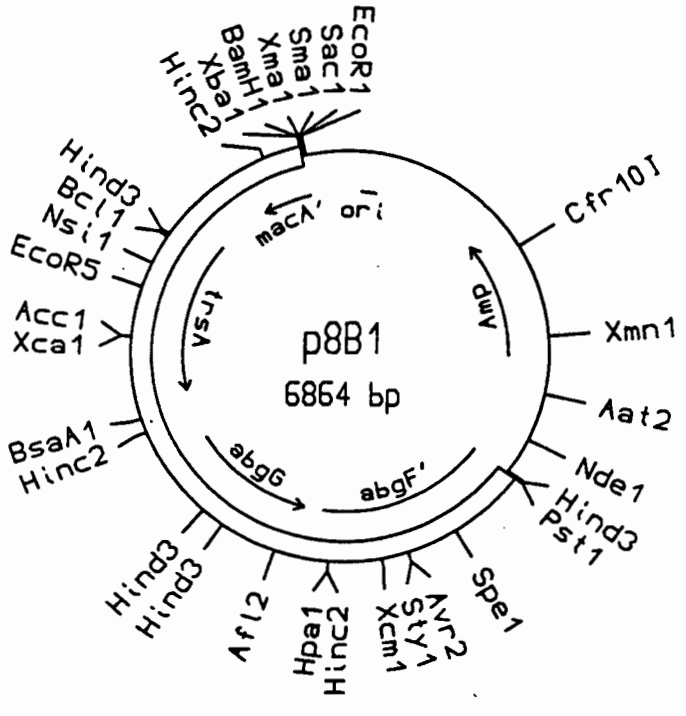


Figure 4.5 Plasmid map of p8B1 showing restriction endonuclease sites and position of the genes on the plasmid. The vector promoter, located 5' to *macA*, is not shown. The open arc represents the *C. longisporum* genomic insert, and the single line represents the vector DNA.

```
          |----> p8B1
1         TCTAGATTTTATGAGCGAGGATGTTAAAAAGCAATTTAATGAATTTATTTTATCTGGTGA
           L  D  F  M  S  E  D  V  K  K  Q  F  N  E  F  I  L  S  G  E

61        ATATTATTATGAAAATGCAGAAAAAATAAGTTCTATATCAGAAGATATTGCTGCTATGTC
           Y  Y  Y  E  N  A  E  K  I  S  S  I  S  E  D  I  A  A  M  S

121       TGAAGAACTTAGTGCATCAATGGAGGAAATAAATTCAATGATTTCAAGTATGGCTGAAAA
           E  E  L  S  A  S  M  E  E  I  N  S  M  I  S  S  M  A  E  N

181       TACTGAAAAGTCAACAGATAACTCTTCTAAGATTTTAGTTAAAGTTAGAGAAACTTCTAA
           T  E  K  S  T  D  N  S  S  K  I  L  V  K  V  R  E  T  S  N

241       TGCTATGGAAGAAGTCGCAGTAACAGCAGAGAAACAGAATGTGTTGGCAAAGGAATTAAA
           A  M  E  E  V  A  V  T  A  E  K  Q  N  V  L  A  K  E  L  N

301       TAAATTAATAAAAGATTTCAAGATATAATAAATACTTTTGAGTTAAATATTGACAACTAG
           K  L  I  K  D  F  K  I  *     ΔmacA                -35

361       GGGTGAAAAACTTAATATGATAATATAAGTTAATAAAGAATAATCTTGTGATAAGGAAGA
                        -10
          ------------------------------------------>          <----------
421       GTAATTAAATTTATTTTCTACACAGAGAGTTAGGAGTATGCTGGAAACCTGACAGAAAAG

          ------------------                                  ------>    <----
481       ATTTAATGAATGGGCCTTTGAAGAATAAAGAGAAATTATTAAAGAGTATCGGATATGGGG
               #
          --                    ---------->          <----------
541       CTCGTGCACCTTATAGCATAGTTAGTATGTTAGTACTAAAACATAGAGAGGTATTTAAAT
                                      #
          ------                ---------->      <----------
601       ACAATTTAGGTGGTACCGCGGAATAAAACTCCGTCCTATTAAAATATAGGATGGAGTTTT
                                    *
661       TTGATGTCTATGAGAATAAAAAAGAGGAGTGGAAAAATATGGCAAAAGAAATTATATTAA
                                  trsA          M  A  K  E  I  I  L  T

721       CAGGTGATAGACCAACAGGAAAATTACATATTGGGCATTATGTAGGTTCATTAAAAAATA
           G  D  R  P  T  G  K  L  H  I  G  H  Y  V  G  S  L  K  N  R

781       GAGTTCAGCTTCAAAATTCAGGAGATTATAGAAGCTTTATAATGATAGCTGATCAGCAGG
           V  Q  L  Q  N  S  G  D  Y  R  S  F  I  M  I  A  D  Q  Q  A

                                                  ---------->         <
841       CTTTAACTGATAATGCAAGAAACCCAGAAAAGATTAGAAATAGTTTGATAGAGGTAGCCT
           L  T  D  N  A  R  N  P  E  K  I  R  N  S  L  I  E  V  A  L

          ----------
901       TAGACTATTTGGCAGTAGGTATTGATCCATTAAAATCTACTATTTTAGTACAATCACAAA
           D  Y  L  A  V  G  I  D  P  L  K  S  T  I  L  V  Q  S  Q  I

961       TACCAGAATTAAATGAACTTACTATGCATTATTTAAATTTAGTTACTCTTTCAAGATTAG
           P  E  L  N  E  L  T  M  H  Y  L  N  L  V  T  L  S  R  L  E

1021      AGCGTAACCCTACAGTAAAGGCAGAAATAAAGCAAAAGAATTTTGAAAATAGTATTCCAG
           R  N  P  T  V  K  A  E  I  K  Q  K  N  F  E  N  S  I  P  A

1081      CAGGATTTTTAATATATCCAGTAAGTCAGGCTGCTGATATCACTGCATTTAAGGCTACTA
           G  F  L  I  Y  P  V  S  Q  A  A  D  I  T  A  F  K  A  T  T

1141      CAGTGCCAGTTGGGGAAGATCAGCTTCCAATGATAGAGCAAGCAAGAGAAATAGTAAGAA
           V  P  V  G  E  D  Q  L  P  M  I  E  Q  A  R  E  I  V  R  S

                                                              ---------->
1201      GTTTTAACACTATATACGGAAAAGAAGTTTTAGTAGAACCAAAGGCTGTTATTCCAAAAG
           F  N  T  I  Y  G  K  E  V  L  V  E  P  K  A  V  I  P  K  G
```

Figure 4.6 continued .....

```
                <----------
1261    GAACTATTGGAAGACTTCCTGGAACAGATGGTAAGGCTAAAATGAGTAAATCTATAGGTA
          T  I  G  R  L  P  G  T  D  G  K  A  K  M  S  K  S  I  G  N

1321    ATGCTATCTATTTAGCTGATGAAGCAGATGTAATAAAACAAAAGGTAATGTCTATGTATA
          A  I  Y  L  A  D  E  A  D  V  I  K  Q  K  V  M  S  M  Y  T

1381    CAGACCCTAATCATATAAAGGTAACAGATCCAGGGCAAGTTGAAGGAAATACTGTATTTA
          D  P  N  H  I  K  V  T  D  P  G  Q  V  E  G  N  T  V  F  T

1441    CCTATTTAGATACCTTCTGCAAAGATACAGAAACATTAGAAGAAATGAAAGCTCATTATA
          Y  L  D  T  F  C  K  D  T  E  T  L  E  E  M  K  A  H  Y  S

1501    GTAGAGGTGGTCTTGGAGATGTTAAGGTTAAAAAATTCTTAAATGAAATATTACAAGCAG
          R  G  G  L  G  D  V  K  V  K  K  F  L  N  E  I  L  Q  A  E

1561    AACTTGAACCTATTAGAAATAGAAGAAAGGAATTTCAAAAAGATATTCCTGAAGTTTATA
          L  E  P  I  R  N  R  R  K  E  F  Q  K  D  I  P  E  V  Y  R

1621    GAATTTTAAAAGAAGGCAGTGAAAAGGCAAGAGAAGTGGCAGCAGGGACTTTAAAAGAAG
          I  L  K  E  G  S  E  K  A  R  E  V  A  A  G  T  L  K  E  V

                                                                  ____
1681    TAAGAGAGACTATTGGTATAGAATATTTTAATAATATATTCTAAAAGGAAGTTTAATAAA
          R  E  T  I  G  I  E  Y  F  N  N  I  F  *

        ---------->     <----------
1741    AATTAGTCTTCATATTAAAGGATATGAAGACTAATTTTTTATGATTATTGATAGAGAGTC

1801    GAAAAAAATCAATATACGTAAATGTAAATTATTAAATTATTAATTTTAATTGTAATTTTT

1861    GATGGTTATATAAAAAACATAAAAAAAAGTTGACATTGTGAATAAAATATAATACTATATA

                                                          _____
1921    CACATAGAAGTTAAACGGTTACATTGATTTCAGATTGTTACTGATTCGATCAGGCAAAAC

        ---------------------->     <----------------------------
1981    CTTAATTAATGAGATGGATTTATATATTATTTATCCGTATTATTAGTTGTTTTTTTTCTT

2041    TTGAAATACTTATAGACTGGGGGGGATATAGTATTTATACCATAAAGAAAATTTTTAATAA
                          abgG          I  Y  T  I  K  K  I  F  N  N

2101    TAATTCAGTTTTAGCTTTAGATTCAGAAAAAAGAGAGATTGTCATTTTAGGATGTGGAAT
          N  S  V  L  A  L  D  S  E  K  R  E  I  V  I  L  G  C  G  I

2161    AGCATTTAAAAAGAAAGTAAATGACAAAGTTTCAGAAGATAACGTTGAAAAAACATTTAT
          A  F  K  K  K  V  N  D  K  V  S  E  D  N  V  E  K  T  F  I

2221    ACTTAAACAAAAGGATGCTTCAGAAAAATTTAAATTATTGTTAGAAGATATTTCAGCAGA
          L  K  Q  K  D  A  S  E  K  F  K  L  L  L  E  D  I  S  A  E

2281    GTATATATCTCTATGTTATGACATAATAGAATATGCTAAAAATATTCTTGATAAAGAGCT
          Y  I  S  L  C  Y  D  I  I  E  Y  A  K  N  I  L  D  K  E  L

2341    TAATGATTACATATATGTAACCCTTACAGATCATATAAATTATGTTATTGAAGCTTGTAA
          N  D  Y  I  Y  V  T  L  T  D  H  I  N  Y  V  I  E  A  C  K

2401    ACAGGGTATATCTAAACCCAATATATTAATATGGGAAATAAAAAAATTCTATCCAAAAGA
          Q  G  I  S  K  P  N  I  L  I  W  E  I  K  K  F  Y  P  K  E

                                                          |-->pGDB2
2461    ATTTGCCGTTGGATTAAAAGCTATAGAGTTTATTGAAGATGAATTAGGGTACAAGCTTCA
          F  A  V  G  L  K  A  I  E  F  I  E  D  E  L  G  Y  K  L  H
```

Figure 4.6 continued .....

```
2521    CGAAGATGAAGCTGGAAATATCGCACTTCATCTAATAAATGCTCAAGTTAATGAAAAATC
         E  D  E  A  G  N  I  A  L  H  L  I  N  A  Q  V  N  E  K  S

2581    TGATACAGTTGAAGATGTATATAATATTACAAAGAAGATAAATGATATTCTTAATATAGT
         D  T  V  E  D  V  Y  N  I  T  K  K  I  N  D  I  L  N  I  V

2641    TAAATATACATACAATCTTGAATTAGATGAAAAAACTTTAAATTATGAAAGATTTATAAC
         K  Y  T  Y  N  L  E  L  D  E  K  T  L  N  Y  E  R  F  I  T

2701    TCATCTAAGATTCTTCTTTAAACGACTTGAAAGAAAAGAAATTACAGAATCTGAAGATAA
         H  L  R  F  F  F  K  R  L  E  R  K  E  I  T  E  S  E  D  N

2761    CTTCTTATTAGAACAAGTAAAAAATAAATATACAGATGCTTATGAATGTATGCTTAAGAT
         F  L  L  E  Q  V  K  N  K  Y  T  D  A  Y  E  C  M  L  K  I

2821    TGAAAAATACTTGGGTCAGGAATTAAGTAATGAAGAACAGTTATATTTAATGTTACATAT
         E  K  Y  L  G  Q  E  L  S  N  E  E  Q  L  Y  L  M  L  H  I

                                                                    ___
2881    TCAACGTGTAACAACAAGAGAAAAGTAAAAAATAGGATTGTTACTGATAATGCAGGCAAG
         Q  R  V  T  T  R  E  K  *

        ------------------------------>      <------------------------------
2941    ACCGTAAAATTTAGAGGAAATTTATATTATAAATTTTGCTTTAAATATTATGTGTCTTGC

3001    CTTTTTATATTAATCATTCTACGAGGAGAGATGAACTATGAAATATGAACAATTAGCTAA
                              abgF          M  K  Y  E  Q  L  A  K

3061    AGACATATTAAAAAATGTCGGTGGCAAAGAAAACGTTAACAGCTTAACACATTGCATTAC
         D  I  L  K  N  V  G  G  K  E  N  V  N  S  L  T  H  C  I  T

3121    TAGATTAAGATTTAAATTAAAAGATGAAGGTAAAGCAAATACAGATATATTAAAAAATAT
         R  L  R  F  K  L  K  D  E  G  K  A  N  T  D  I  L  K  N  M

3181    GGATGGAGTAGTTACCGTAGTTAAAAGCGGTGGACAATATCAAGTAGTTATAGGAAATCA
         D  G  V  V  T  V  V  K  S  G  G  Q  Y  Q  V  V  I  G  N  H

3241    TGTTCCAGACGTATTTGCAGCAGTTAATGCTGTAGGGGGACTTCAAGGAGCTTCAGAAGA
         V  P  D  V  F  A  A  V  N  A  V  G  G  L  Q  G  A  S  E  E

3301    AGATGAAACTGAAGAAAAGATGAATCCTTTAAATAAATTTATAGATATTGTTTCTGGTGT
         D  E  T  E  E  K  M  N  P  L  N  K  F  I  D  I  V  S  G  V

3361    ATTCCAACCAATACTTGGAGTTTTATGTGCTACAGGTATGATTAAAGGTTTAAATGCAGT
         F  Q  P  I  L  G  V  L  C  A  T  G  M  I  K  G  L  N  A  V

3421    ATTAGTTGCAGCAGGAGTTTTAGCTAGTACAGATTCTACTTACATAATACTAAATGCTAT
         L  V  A  A  G  V  L  A  S  T  D  S  T  Y  I  I  L  N  A  I

3481    TGGTGATTGTTTATTTAACTTCTTCCCAATATTCCTAGGTTTTACAGCAGCTAAAAAGTT
         G  D  C  L  F  N  F  F  P  I  F  L  G  F  T  A  A  K  K  F

3541    TAAGTTAAACCAATTTACAGGTATGGCAGTTGGAGCAGCTATGGTTTATCCAGCAATAAC
         K  L  N  Q  F  T  G  M  A  V  G  A  A  M  V  Y  P  A  I  T

3601    TTCATTTGCAGGTCAAACAGTTGATTTCTTTGGTATTCCAGTAGTAATGCCTTCAAGCAG
         S  F  A  G  Q  T  V  D  F  F  G  I  P  V  V  M  P  S  S  S

3661    TTATCAATCAACAGTTATTCCTATAATACTAGCTATATATATAGCATCAAAGGTTGAAAA
         Y  Q  S  T  V  I  P  I  I  L  A  I  Y  I  A  S  K  V  E  K

3721    ATTATTTAAGAAGATTATACCTGATATGGTTAAAACATTCTTAGTTCCATTTGCAACATT
         L  F  K  K  I  I  P  D  M  V  K  T  F  L  V  P  F  A  T  L
```

Figure 4.6 continued .....

```
3781    ACTAGTAGTAGTTCCAGTTACTTTTATGGCAATAGGTCCCATATCAACTATAGCAGCTAA
          L  V  V  V  P  V  T  F  M  A  I  G  P  I  S  T  I  A  A  N

3841    CGCTTTAGGAGATTTAACATTAGCAATTTATAACTTTAATCCAACAATCGCAGGTTTATT
          A  L  G  D  L  T  L  A  I  Y  N  F  N  P  T  I  A  G  L  F

3901    TATTGGTGGATTCTGGCAAGTATTCGTTATGTTTGGTTTACACTGGGGATTAGTTCCAAT
          I  G  G  F  W  Q  V  F  V  M  F  G  L  H  W  G  L  V  P  I

3961    AGCAATGAACAACTTAGCAGTAATTGGTTATGATCCGGTTTTAGCAACTTCAGTTGCAGT
          A  M  N  N  L  A  V  I  G  Y  D  P  V  L  A  T  S  V  A  V

4021    ATGTTTCGCTCAAACAGGTGTTGTAATGGCAATACTTGCAAAAACTAAAGATAAAAAGTT
          C  F  A  Q  T  G  V  V  M  A  I  L  A  K  T  K  D  K  K  L

4081    AAAGTCACTATGTATTCCAGCAATTATTTCAGGATTCTTTGGAGTTACAGAACCAGCTAT
          K  S  L  C  I  P  A  I  I  S  G  F  F  G  V  T  E  P  A  I

4141    CTACGGTATTACATTACCAAGAAAGAAACCATTTATATTAAGCTGTATTGCTGCAGGTGT
          Y  G  I  T  L  P  R  K  K  P  F  I  L  S  C  I  A  A  G  V

4201    TACTGGTGGAATCATAGGATTCTTTGAAAGTAAAGGTTATTCAATGGGTGGTCTTGGAAT
          T  G  G  I  I  G  F  F  E  S  K  G  Y  S  M  G  G  L  G  I

4261    ATTTGCTCTTCCAAGCTATATAAATCCAGAAGGAATTGATAGAGGTTTCTATGGTATGGT
          F  A  L  P  S  Y  I  N  P  E  G  I  D  R  G  F  Y  G  M  V

4321    AATAGCTATGGTAGCTGGTATAGTAGTAGGATTCATATTAATGTTTGTAACTAAACTTAA
          I  A  M  V  A  G  I  V  V  G  F  I  L  M  F  V  T  K  L  N

4381    TGATGAAGAAGAAGTAAAGACAACAGAAAGTAAAAAAGAAGAATCATTAGTTAAACAAGA
          D  E  E  E  V  K  T  T  E  S  K  K  E  E  S  L  V  K  Q  E

4441    AGAAATAGTAAGCCCAATTCAAGGTGAAGTAGTTACTTTAGCAGAAGTTAAAGATGAAGC
          E  I  V  S  P  I  Q  G  E  V  V  T  L  A  E  V  K  D  E  A

4501    ATTCTCATCAGGAGCTTTAGGAAAAGGTGTTGCAATCAACCCTATAGAAGGAAAAGTTTA
          F  S  S  G  A  L  G  K  G  V  A  I  N  P  I  E  G  K  V  Y

4561    TGCACCAGCTGATGGAACATTAACAACTTTATTCCCATCACTACATGCACTTGGTATAAC
          A  P  A  D  G  T  L  T  T  L  F  P  S  L  H  A  L  G  I  T

4621    TACAGAAAATGGTGCAGAAATATTAATACACGTTGGTATGGACACTGTTCAATTAGAAGG
          T  E  N  G  A  E  I  L  I  H  V  G  M  D  T  V  Q  L  E  G

4681    AAAACACTTTACTGCAAAAGTTAAACAAGGTGATAAGATCAAAAAGGGACAATTATTAAT
          K  H  F  T  A  K  V  K  Q  G  D  K  I  K  K  G  Q  L  L  I

4741    AGAATTTGATAAAGAAGCAATAGAAAAAGCAGGGTATTCAACAATTACTCCAGTGCTAAT
          E  F  D  K  E  A  I  E  K  A  G  Y  S  T  I  T  P  V  L  I

4801    TACAAATTCAGATCAATATTTAGATGTAATAGAAACTGATAAAAGAAAAGTAGATGTTAA
          T  N  S  D  Q  Y  L  D  V  I  E  T  D  K  R  K  V  D  V  N

4861    TTCAGAATTATTAACAGTAGTAATATAAGACAGTTAAAGAAAGGCGGAATTTAAAATGTC
          S  E  L  L  T  V  V  I  *                 abgA          M  S

4921    ATTTGATTTAAGTTTTCCAAAAAGTTTTTTATGGGGAGGAGCTACAGCTGCAAACCAATT
          F  D  L  S  F  P  K  S  F  L  W  G  G  A  T  A  A  N  Q  F

4981    TGAGGGAGCGTATAATGAAGATGGAAAGGGCTTATCTATACAAGATATAGCACCAAAAGG
          E  G  A  Y  N  E  D  G  K  G  L  S  I  Q  D  I  A  P  K  G

5041    AGTAATGGGACCTATAACAGAGGTCCCAACAGAAGATAATATGAAACTTATTGGGATAGA
          V  M  G  P  I  T  E  V  P  T  E  D  N  M  K  L  I  G  I  D
```

Figure 4.6 continued .....

```
5101    TTTTTATCACAGATATAAAGAAGATATAAAGCTTTTTGCAGAAATGGGCTTTAAAACTTT
          F  Y  H  R  Y  K  E  D  I  K  L  F  A  E  M  G  F  K  T  F

5161    TAGATTATCTATAGCTTGGTCAAGAATATTCCCAAATGGTGATGATGAGATACCTAATGA
          R  L  S  I  A  W  S  R  I  F  P  N  G  D  D  E  I  P  N  E

5221    AAAAGGTTTGGAATTTTATGACAAAGTATTTGATGAACTTCAAAGGTATGGAATTGAACC
          K  G  L  E  F  Y  D  K  V  F  D  E  L  Q  R  Y  G  I  E  P

5281    ACTAGTAACTTTATCTCACTATGAAACTCCACTTAATTTATCTAAGAAATATAATGGATG
          L  V  T  L  S  H  Y  E  T  P  L  N  L  S  K  K  Y  N  G  W

5341    GGCAAATAGAGATCTTATAGGCTTCTATGAAAGATATGTAAGAACTGTATTCACTAGATA
          A  N  R  D  L  I  G  F  Y  E  R  Y  V  R  T  V  F  T  R  Y

5401    TAAAGATAAGGTAAAATACTGGCTTACTTTTAATGAAATAAACTCAGCAATACATGCTCC
          K  D  K  V  K  Y  W  L  T  F  N  E  I  N  S  A  I  H  A  P

5461    ATATATGAGTGCAGGAATTTGGACTGATAAGTCAGAATTAAGTAAACAAGATTTATATCA
          Y  M  S  A  G  I  W  T  D  K  S  E  L  S  K  Q  D  L  Y  Q

5521    AGCAATGCATCATGAATTAGTTGCAAGCGCCTTAGCAGTAAAAATTGGACATGAAATTAA
          A  M  H  H  E  L  V  A  S  A  L  A  V  K  I  G  H  E  I  N

5581    CCCTGATTTTAAAATAGGATGTATGATTCTTGGAATTCCTGTATATCCATTAACACCACA
          P  D  F  K  I  G  C  M  I  L  G  I  P  V  Y  P  L  T  P  H

5641    TCCTGATGATTTAATTGAAAAAATGAGAGTGGAAAGAGAAAGTCTATTCTTTGCAGATGT
          P  D  D  L  I  E  K  M  R  V  E  R  E  S  L  F  F  A  D  V

5700    ACATGCTAGAGGAAAATATCCAAGATATATGAATAGATTATTTAAGGAAAATAATATAGA
          H  A  R  G  K  Y  P  R  Y  M  N  R  L  F  K  E  N  N  I  E

5761    AATAAAATGGCATGAAGATGATGCAGAGATTTTAAGTAATGTAGTAGATTTTATATCATT
          I  K  W  H  E  D  D  A  E  I  L  S  N  V  V  D  F  I  S  F

5821    TAGCTATTACATGAGCTCATGTGCAACAGCAGATGAAGAAAAGAAGAAAGCAGGAGCTGG
          S  Y  Y  M  S  S  C  A  T  A  D  E  E  K  K  K  A  G  A  G

5881    CAATTTATTAGCTGGAGTACCTAATCCATATTTAAAAGCATCTGAATGGGGATGGCAAAT
          N  L  L  A  G  V  P  N  P  Y  L  K  A  S  E  W  G  W  Q  I

5941    TGATCCTAAGGGACTTAGATTAATACTTAATGAATTATATGATAGATATGAAAAGCCATT
          D  P  K  G  L  R  L  I  L  N  E  L  Y  D  R  Y  E  K  P  L

6001    ATTTATAGTTGAAAATGGTCTTGGAGCTGTAGATGAACTTGTTACTGATGAAAATGGTAA
          F  I  V  E  N  G  L  G  A  V  D  E  L  V  T  D  E  N  G  N

6061    TAAAACAGTAAATGATGACTATAGAATTAAGTATTTAAATGACCATCTTGTTCAAGTAGC
          K  T  V  N  D  D  Y  R  I  K  Y  L  N  D  H  L  V  Q  V  A

6121    AGAAGCTATAGAAGATGGAGTTGAACTTATGGGATATACAACTTGGGGATGTATTGACTT
          E  A  I  E  D  G  V  E  L  M  G  Y  T  T  W  G  C  I  D  L

6181    AGTTTCAGCTTCAACAGCAGAACTGAAGAAGAGATATGGATTTATCTATGTTGATAGACA
          V  S  A  S  T  A  E  L  K  K  R  Y  G  F  I  Y  V  D  R  H

6241    TGATGATGGCTCAGGAACTTTAGAAAGATATAAAAAGAAGAGCTTTAATTGGTACAAAGA
          D  D  G  S  G  T  L  E  R  Y  K  K  K  S  F  N  W  Y  K  E

                                                          ----------

6301    AGTAATTGCTACTAATGGAAAGAGCTTAGAAAGATAATTATTAATAAATATAGAAATGAA
          V  I  A  T  N  G  K  S  L  E  R  *
```

Figure 4.6 continued .....

```
                 ---->                <----------
6361   TCTGAAAAGTTAAAAATTCAGATTCATTTCTTTTTGTGTTTAGAAAAGATATAAAGATGA

6421   AATAATTAGAAACTAGGAGGGATTTATGGATATAGAAAAAGTTATAAAATTTATAGATAA
                       ORF6      M  D  I  E  K  V  I  K  F  I  D  K
                                                  ---------->
6481   AAGTATAGATGAAAAAGAAAATTATAATGAGCTTATTATAAAGGAATTAATTAAAGATAA
        S  I  D  E  K  E  N  Y  N  E  L  I  I  K  E  L  I  K  D  K
               <----------
6541   GAATGGGTTAATTCCTCTTTTAGTTAAAATTGATATATTTTATTTTTTGTATATACTCAA
        N  G  L  I  P  L  L  V  K  I  D  I  F  Y  F  L  Y  I  L  K

6601   AGTTTTAGATGAAGTGATTTTAATACTATATAGTAAAGATTTTATAAGTGAGTTTAAAAG
        V  L  D  E  V  I  L  I  L  Y  S  K  D  F  I  S  E  F  K  R

6661   AAAAATATACATAAGGTATGATAAACTTTCAAATATAAATGTAAATAACCATGAAACAAT
        K  I  Y  I  R  Y  D  K  L  S  N  I  N  V  N  N  H  E  T  I

6721   AAAAGATAAATGTAATAATTAATATATAAAGGTTTTTGTAATGATTTGAAAATATAACTA
        K  D  K  C  N  N  *

6781   AGTTAAATCATATAAATAAGCTATAATATAAAGTAAGAAAATTAAATAAGATTTATGGGA

6841   TTAGGAAGTGTATCGTTATGAAGAAAATTGAAGCTGTAATACGTCCGGAAAAACTTGAAG
           glnB          M  K  K  I  E  A  V  I  R  P  E  K  L  E  E

6900   AATTAAAGGATGCTTTAAAAAAAGGCAATATAAACGGAATTACTGTAGATCAAGTAATGG
        L  K  D  A  L  K  K  G  N  I  N  G  I  T  V  D  Q  V  M  G

6961   GATGTGGTCAGCAACATGGATGGACTGAACATTATAGAGGAAGCGAAATCATGGTGAACT
        C  G  Q  Q  H  G  W  T  E  H  Y  R  G  S  E  I  M  V  N  F

7021   TTTTACCTAAGGTTGAAGTAAAGGTTATAGTTGATGACTCAAAGGTGCAAGAGGTCATAA
        L  P  K  V  E  V  K  V  I  V  D  D  S  K  V  Q  E  V  I  N

7081   ATTTAATTATTGATACAGTAAAAACAGGGGAAGTTGGAGACGGAAAAATATTTGTTAGTG
        L  I  I  D  T  V  K  T  G  E  V  G  D  G  K  I  F  V  S  D

7141   ATGTTAGTGAATGCATAAGAATAAGGACAGGAGAAAAAGGATTAGATGCATTATAAAAGA
        V  S  E  C  I  R  I  R  T  G  E  K  G  L  D  A  L  *

                    ---------->          <----------
7201   ATTTATTATTAAAAAATATTACTGGAAATTTAGTTTTCCAGTAATATTTTTTTAAAGTCT

7261   AAATTTTTTCTTTTTATACTTTTTGAAAGTTTTAAATAGTGCAGATATAAATATTATCCC

7321   AAGAGCTAGTGCACCAGCATTGTATAAAGTAGTAACTCCATAATCTATGGCACTAATATA

7381   ATCTCCAACTAAGAAAGCATACATACATTTATATACCCCATATCCAGGAACAACAGGAAT

7441   AAGAGCACACACACTTAATAAAGTTACAGGGGTATTAAATTTTCTTGCACAAATTTCACT

7501   ATATAAACTAAAGGCTGC
```

Figure 4.6 The complete 7.5 kb DNA sequence derived from genomic inserts of *C. longisporum*. Sequence was obtained from pGDB2, p4.9, p4.2 and p8B1 as described. Gene names are presented underneath the proposed ribosome binding sites (bold) except for *macA* which is at the end of the gene. Single underlined sequence indicate -10 and -35 regions. Overline arrows indicate inverted repeats. Double underlined sequence identifies Box A/ Box B motifs (see text). Asterisks indicate stop codons. The start of p8B1 and pGDB2 are indicated. Amino acids are represented by single letters at the first base of each codon. The start codons of the putative 5' ORF of *trsA* are indicated by #.

Table 4.2 Percent protein sequence identity of each ORF on pGDB2 and p8B1 to known sequences as indicated (referenced in the text). Numbers in brackets represent percent similarity. NCBI accession numbers are given for unpublished genes.

Comparison sequence	*ΔmacA*	*trsA*	*abgG*	*abgF*	*abgA*	*glnB*
B. subtilis TlpB C-terminus TlpC C-terminus McpA C-terminus	27(45) 28(46) 32(51)					
B. stearothermophilus TrsS		33(54)				
E. coli TrsS		33(54)				
C. elegans C34E100.4		34(54)				
B. subtilis NCBI (534045)			47(67)			
E. coli BglG			36(62)			
E. chrysanthemi ArbG			42(63)			
B. subtilis SacY SacT			36(59) 38(57)			
E. coli BglF AscF				44(66) 30(56)		
E. chrysanthemi ArbF				42(67)		
B. lactofermentum PtsG				33(60)		
E. coli AscB BglB					55(71) 60(74)	
E. chrysanthemi ArbB					61(74)	
Synechococcus P_{II}						56(74)
A. braseilense P_{II}						55(71)
B. subtilis NrgB						42(60)

4.4.3 Analysis and comparison of each ORF

macA: This truncated ORF, containing 108 C-terminal amino acids (from bp 1 to 325), is similar to the C-terminal regions of a number of transmembrane methyl-accepting chemotaxis proteins from *B. subtilis* (84) (Figure 4.7). Although translation presumably ends at the UAA codon, no transcription terminators were identified.

In *E. coli*, methyl-accepting chemotaxis proteins (MCPs) recognise either ligand-occupied forms of periplasmic binding proteins or the molecules themselves. When these effector molecules bind to the receptor they induce conformational changes, transferring the information into the cytoplasm by affecting the activity of a cytoplasmic autophosphorylating protein kinase (CheA). Phospho-CheA is central to the mechanisms involved in excitation, which is the immediate response to attractant or repellent stimuli, and adaption. In a series of excitatory reactions involving phosphoryl transfer the swimming behaviour of the bacterium is altered (124,211). The adaption to the stimulus is controlled by methylation of glutamate residues on the C-terminal region of the receptor. The MCPs of *B. subtilis* serve similar functions but are not the direct receptors for the binding of chemoeffectors (84,211).

```
TlpB        ISSGDLTEVIDIHSKNEFGQLGESFNEMSASLRSVIGVIQTSVENVASSSEELTASAAQT
McpA        ISEGDLTETIDIRSKDELGELGKSFNNMASSLRSLIHAIQDSVDNVAASSEELTASAAQT
MacA        ---------LDFMSED-----------------------------------------
                      .*  *

TlpB        SKATEHITLAIEQFSDGNEAQSEKLETSSNHLSQMNEGISKVAQASSTITKSSIQSSEAA
McpA        SKATEHITLAIEQFSNGNEKQNENIETAAEHIYQMNDGLTNMAQASEVITDSSVQSTEIA
MacA        ---------VKKQFN-----------------------------------------
                       **

TlpB        GSGEKLVEHTVGQMKTIDQSVQKAEAVVKGLETKSQDITSILNVINGIADQTNLLALNAA
McpA        SEGGKLVHQTVGQMNVIDKSVKEAEQVVRGLETKSKDITNILRVINGIADQTNLLALNAA
MacA        ------------------------------------------------------------

TlpB        IEAARAGEYGRGFSVVAEEVRKLAVQSADSAKEIEGLIQEIVREISTSLSMFQSVNHEVK
McpA        IEAARAGEYGRGFSVVAEEVRKLAVQSADSAKEIEGLIIEIVKEINTSLGMFQSVNQEVQ
MacA        -EFILSGEYY----------------------------------------------------
             *   .***

                                                 ▼
TlpB        EGLQITDQTAESFKQIYEMTTQISGELQNLN ATVEQLSAG SQEVSSAVEDISAVAKESSA
McpA        TGLDITDKTEMSFKRISEMTNQIAGELQNMS ATVQQLSAS SEEVSGASEHIASISKESSA
MacA        ----------------YENAEKISSISEDIA AMSEELSAS MEEINSMISSMAENTEKSTD
                             * . .*.   . . *  ..***  .*.      ..  .  *.

                    ▼     ▼
TlpB        GIQDIAAS AEEQLASMEEISSS AETLANMAEELQDITKKFKIES
McpA        HIQDIAAS AEEQLASMEEISSS AETLSSMAEELRDMTKRFKIE--
MacA        NSSKILVKVRETS NAMEEVAVT AEKQNVLAKELNKLIKDFKI---
               *     *   .***.. .**    .* **  . * ***
```

Figure 4.7 Amino acid alignment of ΔMacA to the C-terminal regions of TlpB and McpA from *B. subtilis*. Asterisks under the sequence indicate identical residues in all three sequences and dots indicate a conservative change. Bold residues indicate methylation site sequences while arrows above the sequence indicate the methylated residue. Dashes show gaps created by the CLUSTALW program (97), which was used to produce the alignment.

Based on the similarities to the *B. subtilis* proteins, the truncated MacA (ΔMacA) appeared to code for part of the cytoplasmic domain. As these C-terminal domains are thought to have conserved structural features (84) putative methylation sites were determined by homology to the *E. coli* consensus sequence and positional relatedness to the *B. subtilis* sites (Table 4.3, Figure 4.7). Interestingly, the putative methylation sites on ΔMacA corresponded in location to those of *B. subtilis*, which differs in location from the *E. coli* MCPs (84).

Table 4.3 Comparison of the *E. coli* consensus methylation sequence, proposed *B. subtilis* MCP methylation sites in McpA and TlpB (84) and proposed ΔMacA methylation sites. Asterisk denotes sites of methylation.

E. coli consensus methylation sequence	* A-*X*-*X*-E-E-*X*-A-A-S (S) (S) (T) (T) (A)
McpA 1	A-S-S-E-E-L-T-A-S
2	A-T-V-Q-Q-L-S-A-S
3	A-S-A-E-E-Q-L-A-S
4	A-S-D-E-E-I-S-S-S
TlpB 1	S-S-S-E-E-L-T-A-S
2	A-T-V-E-Q-L-S-A-G
3	A-S-A-E-E-Q-L-A-S
4	A-S-M-E-E-I-S-S-S
ΔMacA 1	A-M-S-E-E-L-S-A-S
2	N-A-M-E-E-V-A-V-T

Future sequence analysis should identify the entire *macA* sequence from p81, which contains uncharacterised upstream sequence presumably containing the missing 5' regions of this gene. This analysis should also include a determination of the various other domains, such as the signaling domain of which a part may be present in ΔMacA (results not shown), and the chemoeffectors.

trsA: Separated by 372 bp from *macA*, *trsA* encodes a polypeptide of 341 amino acids (from bp 699 to 1721; 37.5 kDa) which is similar to tryptophanyl-tRNA synthetase genes from *E. coli* (82,83), *Bacillus stearothermophilus* (20) and the nematode, *Caenorhabditis elegans* (224) (Figure 4.8; Table 4.2). The products of these genes are responsible for activating amino acids and transferring them to their respective tRNA molecules (186).

The transcribed gene is presumably translated from the AUG start codon and ends at a UAA stop codon. The gene is preceded 7 bp upstream by a sequence (GAGGAGUGG) which may function as a ribosome binding site (RBS) (193). A -10 and -35 region, upstream of the approximate 250 bp leader sequence (see below), were tentatively identified. Transcription of *trsA* is presumably terminated by the *rho*-independent terminator located at the 3' end of this gene (ΔG -17.9). Interestingly, two internal regions of dyad symmetry were present, which could form hairpin structures (ΔG -5.2 and -6.6), but their potential function(s) are unknown (Figure 4.6).

In class I aminoacyl-tRNA synthetases, which include synthetases specific for tryptophan, two specific conserved sequences have been identified in the coding region (140,187,221). TrsA possesses both of these conserved sequences including the HIGH motif (amino acids 22-25) in the N-terminal segment, and the KMSKS motif (amino acids 201-206) in the C-terminal region (Figure 4.8). The HIGH region has been shown to be part of the adenylate binding site, involved in amino acid activation, while the KMSKS region is thought to be involved in binding the acceptor arm of the tRNA molecules and/or the transfer of the amino-acid onto the tRNAs (32,140,187). The presence of these two highly conserved regions in TrsA suggests that it is a Class I tRNA synthetase.

The leader region of the *trsA* gene also shares structural features and conserved elements with a group of Gram positive aminoacyl-tRNA synthetase and amino acid biosynthesis genes, termed the T-box genes (Figure 4.9). These conserved elements include three stem loop structures which precede a highly

conserved 14-bp element (the T-box), and a *rho*-independent terminator (78,79,91,92). T-box genes are regulated by a common mechanism of transcription antitermination, mediated by the levels of charged and uncharged tRNA molecules in the cell (79,91). Uncharged tRNA molecules bind to a specific acceptor codon, for the appropriate amino acid, on the side bulge of stem I while the acceptor arm of the tRNA binds to a specific sequence (within the T-box) and

```
BS TrpS          ------------------MKTIFSGIQPSGVITIGNYIGALRQFVELQHEYN----CYF
EC TrpS          -----------------MTKPIVFSGAQPSGELTIGNYMGALRQWVKMQDDYH----CIY
CE C34E100.4     MRNYDFQPNNLRLLSTSTHPTIYFTGIQPTGIPHLGNFFGSIEPWTELQNSVDKNILMML
CL TrsA          -----------------MAKEIILTGDRPTGKLHIGHYVGSLKNRVQLQNSGDYRSFIM-
                                      .* .*.*   .*.. *..     .*

BS TrpS          CIVDQHAITV-WQDPHELRQNIRRLAAKYLAVGIDPTQATLFIQSEVPAHAQAAWMLQCI
EC TrpS          CIVDQHAITV-RQDAQKLRKATLDTLALYLACGIDPEKSTIFVQSHVPEHAQLGWALNCY
CE C34E100.4     SVVDQHAISLGPLPANELRQNTHQMTASLIACGVDPNRTLLFRQSDVPQIAQISWILGSL
CL TrsA          -IADQQALTDNARNPEKIRNSLIEVALDYLAVGIDPLKSTILVQSQIPELNELTMHYLNL
                  . **.*..        .*          .* *.** .. .  ** .*   .

BS TrpS          VYIGELERMTQFK-EKSAG--KEAVSAGLLTYPPLMAADILLYNTDIVPVGEDQKQHIEL
EC TrpS          TYFGELSRMTQFK-DKSARY-AENINAGLFDYPVLMAADILLYQTNLVPVGEDQKQHLEL
CE C34E100.4     QTTSKLARLPQYK-EKKERFKKGDIPVGLLTYPLLQAADVLTFKATTVPVGEDQSQHLNL
CL TrsA          VTLSRLERNPTVKAEIKQKNFENSIPAGFLIYPVSQAADITAFKATTVPVGEDQLPMIEQ
                      * *   * .         .  *   **   ***.  . .  *******   .

BS TrpS          TRDLAERFNKRYG-ELFTIPEARIPKVGARIMSLVDPTKKMSKSDPNPKAYITLLDDAKT
EC TrpS          SRDIAQRFNALYG-EIFKVPEPFIPKSGARVMSLLEPTKKMSKSDDNRNNVIGLLEDPKS
CE C34E100.4     LGGLAYAFNKTYETEIFPIPKQLTRESHARIRSLREPEKKMSKSSGGPRSRIEITDSRST
CL TrsA          AREIVRSFNTIYGKEVLVEPKAVIPKGTIGRLPGTDGKAKMSKSIGNA---IYLADEADV
                    .   **  *  *.   *                .   *****  .    * . .

BS TrpS          IEKKIKSAVTDSEGT--IRYDKEAKPGISNLLNIYSTLSGQSIEELERQYEGKGYGVFKA
EC TrpS          VVKKIKRAVTDSDEPPVVRYDVQNKAGVSNLLDILSAVTGQSIPELEKQFEGKMYGHLKG
CE C34E100.4     IIEKCQKA--QSDNAGKVTYDKENRLAVSNLLDLYSAVTKTQTSEID--FSNWTTLDLKM
CL TrsA          IKQKVMSMYTDPNHIKVTDPGQVEGNTVFTYLDTFCKDTET-LEEMKAHYSRGGLGDVKV
                 .  *                       . . *   .  .    *.   .        *

BS TrpS          D--LAQVVIETLRPIQERYHHWMESEE-LDRVLDEGAEKANRVASEMVRKMEQAMGLGRR
EC TrpS          E--VADAVSGMLTELQERYHRFRNDEAFLQQVMKDGAEKASAHASRTLKAVYQAIGFVAK
CE C34E100.4     N--LAEAVDKRLAPIRQKFEELQNTGE-VDKVLTENGEKAREIAEKNLEEIRRTIGFL--
CL TrsA          KKFLNEILQAELEPIRNRRKEFQKDIPEVYRILKEGSEKAREVAAGTLKEVRETIGIEYF
                    .   .   *  .. .          . ... .. ***   *   .  .  ..*

BS TrpS          R----
EC TrpS          R----
CE C34E100.4     -----
CL TrsA          NNIF-
```

Figure 4.8 Sequence alignment of TrsA to tryptophanyl tRNA synthetases from *B. stearothermophilus* (BS), *E. coli* (EC), *C. elegans* (CE) and *C. longisporum* (CL). Asterisks below the sequence indicate identical residues while dots represent a conservative change. Bold lettering indicates conserved sequences mentioned in the text. CLUSTALW (97) was used to produce the alignment.

stabilises the formation of an antiterminator, allowing transcription. Conversely, charged tRNAs cannot bind to and stabilise the antiterminator, resulting in the formation of the *rho*-independent terminator which thereby prevents transcription (Figure 4.10) (78,79,91). The $tRNA^{trp}$ acceptor codon (TGG) on a side bulge of stem I, the T-box, a putative antiterminator structure, and other conserved regions were identified in the leader region of *trsA* (Figure 4.10).

The leader sequence also contains a putative short ORF that could begin at the AUG codon at positions 486-489 or 566-569 and end at the UAA stop codon (624-626), although no putative RBS was identified (Figure 4.6). The translation of short leader peptides have been associated with transcription attenuation in a number of other amino acid and aminoacyl-tRNA synthetase genes, such as the *trsS* gene of *E. coli* (83,91,165,231). In these systems, the leader ORF normally contains several codons specifying the appropriate amino acid, a feature not detected in the putative leader peptide of *trsA*. Furthermore, in T-box genes so far described, there is no conservation of the position or length of putative leader region ORFs and no obvious translational start signals (91). The lack of putative translational start signals and specifier codons for the leader peptide further suggests that *trsA* is a T-box Class I tRNA synthetase, which is not regulated by translational attenuation of transcription.

Any future work on this gene, and its product, should address two main issues. Firstly what role, if any, the regions of dyad symmetry within the coding region play and, secondly, whether the potential 5' leader ORF is translated and/ or involved in regulation, or is just a chance feature. It would also be interesting to determine if all the amino acid biosynthetic and aminoacyl-tRNA synthetase genes of this organism, and other rumen Gram positives, are T-box genes as suggested by Grundy and Henkin (1994).

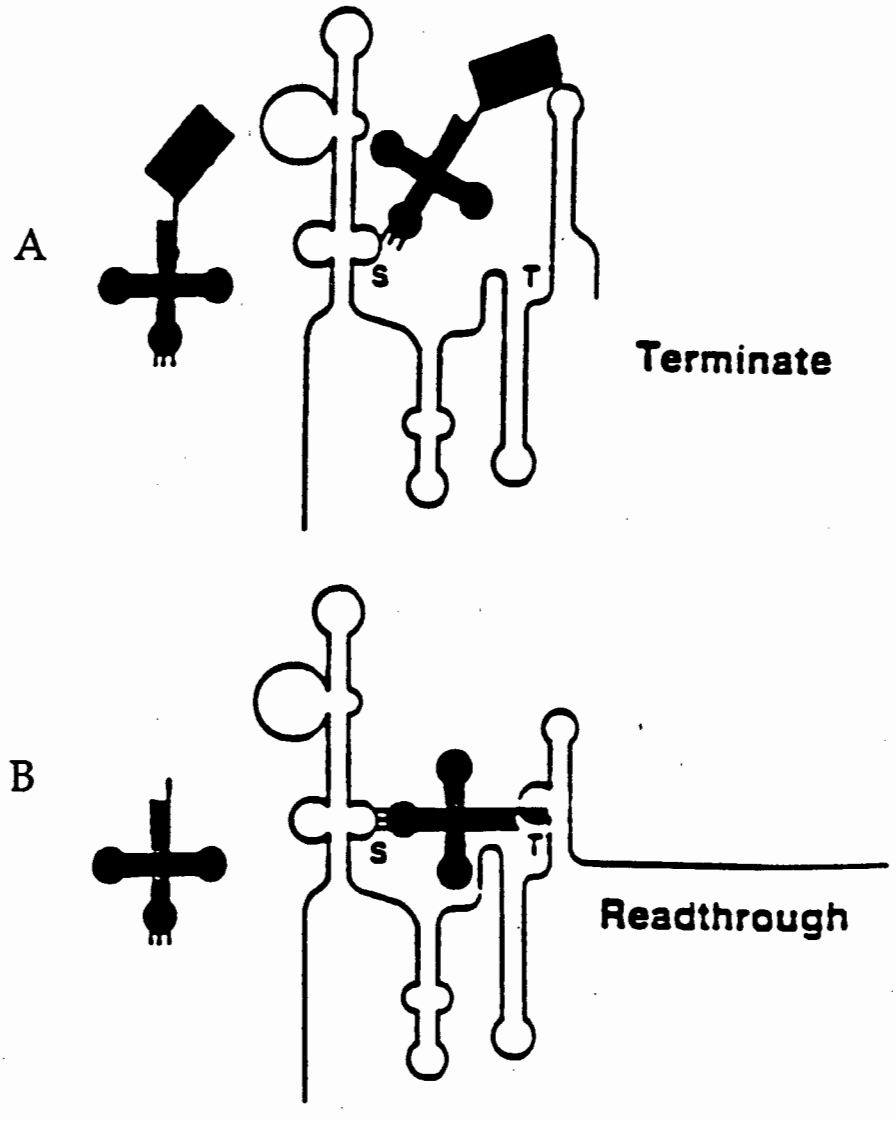


Figure 4.9 Model for induction of T-box tRNA synthetase genes. The tRNA molecule, charged or uncharged with an amino acid, is represented by a cloverleaf structure. (A) represents the terminator form, while (B) represents the antiterminator form of the leader of T-box genes. Also indicated is the specifier sequence (S) and T-box (T). Uncharged tRNA interacts with the specifier sequence and T-box stabilising the antiterminator structure and allowing continued transcription. Conversly, charged tRNA cannot interact with both the specifier and T-box allowing terminator formation and termination of transcription (from (91)).

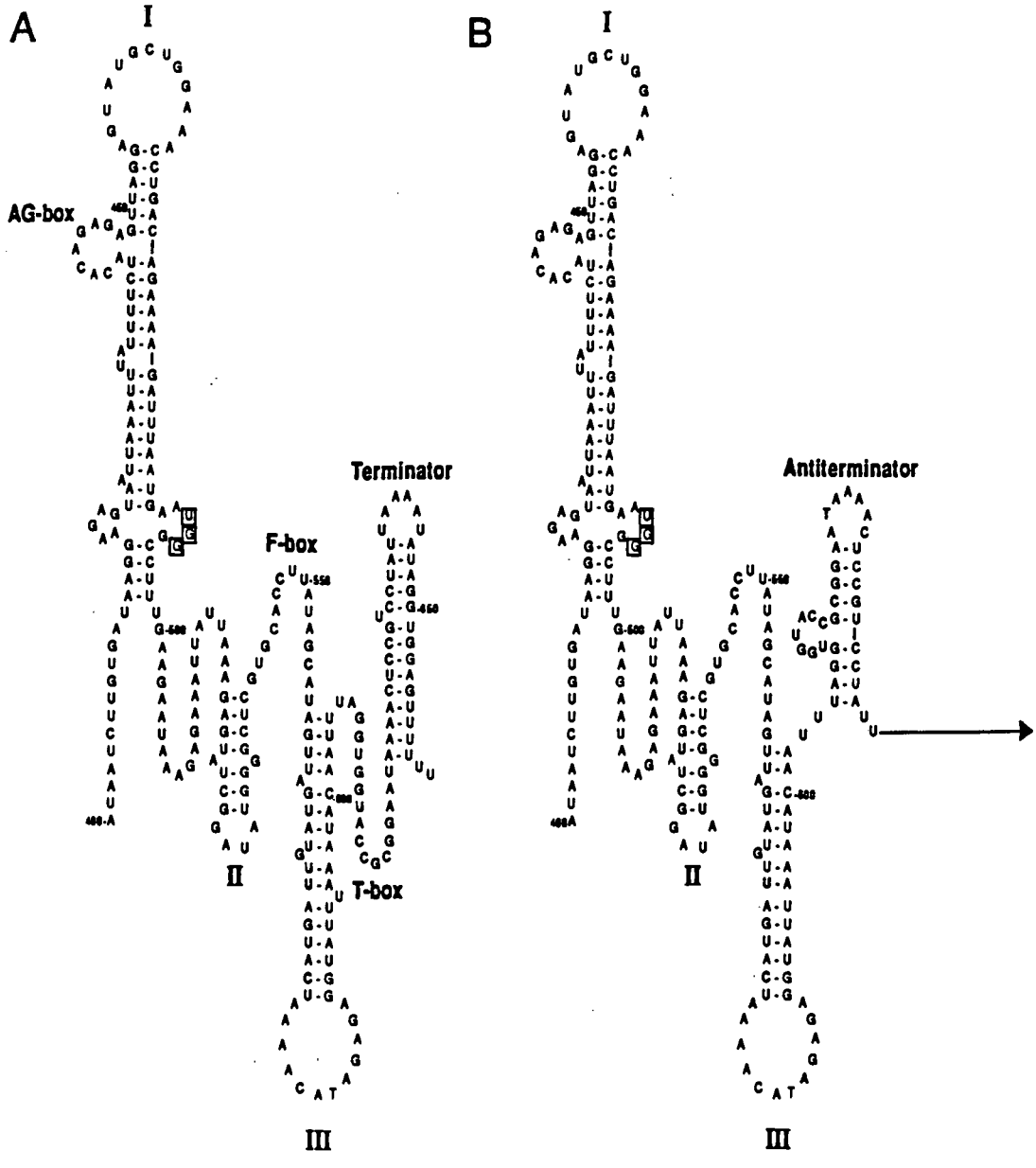


Figure 4.10 Structural model of *C. longisporum trsA* leader. Numbers indicate sequence position. Roman numerals indicate stem-loops I, II and III. The boxed residues are the UGG specifier sequence. AG- box, F-box and T-box are conserved primary sequence elements. Drawn according to Grundy and Henkin (1993).
A. Terminator form
B. Antiterminator form. The arrow indicates continued transcription.

abgG: Separated from *trsA* by 351 bp, this ORF encodes a protein 278 amino acids long (from bp 2072 to bp 2905; 30.6 kDa) and is similar to proteins involved in the antitermination of sugar utilisation genes and operons from *B. subtilis* (45,204,239), *E. chrysanthemi* (85,86) and *E. coli* (192) (Figure 4.11). The putative AUU start codon (171) of this protein is rare, and unverified, but was derived from homology analysis, the general length of these proteins (277 - 283 residues) and from the position of the putative ribosome binding site (UGGGGGGA; Figure 4.6). Translation of AbgG presumably ends at the UAA stop codon (bp 2906-2908). No -10 or -35 regions were identified.

In *E. coli*, the antiterminator protein (BglG) plays a role in the regulation of the *bgl* operon, which contains genes involved in the uptake and utilisation of aromatic-ß-glucosides (including salicin and arbutin) (192). This operon is normally cryptic in *E. coli* and is activated by mutations such as the transposition of insertion sequences (IS) into the promoter region (161,190). Dimerised BglG recognises specific sequences, termed Box A/ Box B or ribonucleic antiterminator (RAT) motifs, which are located preceding and extending into each terminator located upstream of both *bglG* and *bglF*. The dimerised protein prevents transcription termination by binding to the nascent RNA and blocking the formation of the terminator structures (see Figure 4.19) (12,13,102,126,127,151,188). SacY and SacT from *B. subtilis*, and ArbG from *E. chrysanthemi* are similar to BglG and are also thought to act as transcriptional antiterminators, acting on similar conserved regions preceding *rho*-independent terminator like structures (17,44,45,85,86,239).

The presence of these conserved regulatory elements upstream of both *abgG* and *abgF* were determined (Figure 4.6, Table 4.4). BoxA / BoxB motifs were detected which preceded and extended into two *rho*-independent terminator structures, upstream of both *abgG* (ΔG -10.1; *abgG* 1) and *abgF* (ΔG -26.3; *abgG* 2). BoxA and BoxB sequences can form weak stem-loop structures and are thought to be the binding sites of BglG (17,102). The binding of BglG to these sequences, which extend into the terminators, would physically block terminator formation

and result in antitermination (13). Furthermore, some of the unpaired bases in the secondary structure control the specificity of the antiterminator interactions (17). The predicted secondary structure of the *abgG* 1 and 2 BoxA /BoxB motifs are shown in Figure 4.12.

The antitermination activity of BglG, in *E. coli*, is negatively controlled by phosphorylation (10,12). The site of phosphorylation on BglG occurs at a conserved region, containing Asp and His residues, and is catalysed by the *bglF* gene product (44,45,86). Phosphorylation is thought to prevent BglG dimerisation, necessary for binding to the Box A/B sequence (12,13). The proposed AbgG protein also contains these conserved residues and in similar positions (Figure 4.11).

Table 4.4 BoxA / BoxB motifs from *B. subtilis* (BS), *E. coli* (EC), *E. chrysanthemi* (ECh) and *C. longisporum* (CL) leader regions of the genes indicated, a proposed consensus sequence, and the first base of the terminator inverted repeat.

Gene	BoxA	BoxB
BS *sacB*	TCGCGCGGGTTTGTTACTGATAAA	GCAGGCAAGACCTAAAATG
BS *sacPA*	ATAAGCGGGATTGTGACTGGTAAA	GCAGGCAAGACCTAAAATT
EC *bglG*	AATGACTGGATTGTTACTGCATTC	GCAGGCAAAACCTGACATA
EC *bglF*	GAGTAAAGGATTGTTACCGCACTAAG	CAGGCAAAACCTGAAAAA
Ech *arb*	CAATCCGGGGTTGCTACTGCCATTGG	CAGGCAAAACCAGATGTT
CL *abgG* 1	TGATTTCAGATTGTTACTGATTCGAT	CAGGCAAAACCTTAATTA
CL *abgG* 2	AAAAATAGGATTGTTACTGATAAT	GCAGGCAAGACCGTAAAAT
Consensus	NNNGGNTTGTTACTGNNNNN A CG C NT	GCAGGCAAAACCNNANNN G
Terminator Start		|--------->

```
SacY            --MKIKRILNHNAI-VVKDQNEEKILLGAGIAFNKKKNDIVDPSKIEKTFIRKDTPDYKQ
SacT            --MKIYKVLNNNAA-LIKEDDQEKIVMGPGIAFQKKKNDLIPMNKVEKIFVVRD--ENEK
BglG            MNMQITKILNNNVVVVIDDQQREKVVMGRGIGFQKRAGERINSSGIEKEYALSSHELNGR
ArbG            --MKIAKILNNNVVTVMDEQNNEQVVMGRGLGFKKRPGDTVNAALIEKIFSLRSSELTAR
BS reg          --MKIAKVINNNVISVVNEQGKELVVMGRGLAFQKKSGDDVDEARIEKVFTLDNKDVSEK
AbgG            -IYTIKKIFNNNSVLALDSEKREIVILGCGIAFKKKVNDKVSEDNVEKTFILKQKDASEK
                    * .. *.*    .      * ...* *. * *. .. .    .** .         .

SacY            FEEILETLPEDHIQISEQIISHAEKELNIKINERIHVAFSDHLSFAIERLSNGMVIKNPL
SacT            FKQILQTLPEEHIEIAEDIISYAEGELAAPLSDHIHIALSDHLSFAIERIQNGLLVQNKL
BglG            LSELLSHIPLEVMATCDRIISLAQERL-GKLQDSIYISLTDHCQFAIKRFQQNVLLPNPL
ArbG            LSDVLERIPLEVVTTADRIIALAKEKLGGNLQNSLYISLTDHCHFAIERHRQGVDIRNGL
BS reg          FKTLLYDIPIECMEVSEEIISYAKLQLGKKLNDSIYVSLTDHINFAIQRNQKGLDIKNAL
AbgG            FKLLLEDISAEYISLCYDIIEYAKNILDKELNDYIYVTLTDHINYVIEACKQGISKPNIL
                   .*  .  . . .  **  *   *   .  . .. .**  . *     ..   * *

SacY            LNEIKVLYPKEFQIGLWARALIKDKLGIHIPDDEIGNIAMHIHTARNNAGD--MTQTLDI
SacT            LHEIKALYKKEYEIGLWAIGHVKETLGVSLPEDEAGYIALHIHTAKMDAES--MYSALKH
BglG            LWDIQRLYPKEFQLGEEALTIIDKRLGVQLPKDEVGFIAMHLVSAQMSGN---MEDVAGV
ArbG            QWEVKRLYQKEFAIGLDALDIIHRRLGVRLPEDEAGFIALHLVNAQLDSH---MPEVMRI
BS reg          LWETKRLYKDEFAIGKEALVMVKNKTGVSLPEDEAGFIALHIVNAELNEE---MPNIINI
AbgG            IWEIKKFYPKEFAVGLKAIEFIEDELGYKLHEDEAGNIALHLINAQVNEKSDTVEDVYNI
                  . .  *  *. .*  *    .    *  .  ** * **.*.  *          .

SacY            TTMIRDIIEIIEIQLSINIVEDTISYERLVTHLRFAIQHIKAGESIYELDAEMIDIIKEK
SacT            TTMIKEMIEKIKQYFNRKVDENSISYQRLVTHLRYAVSRLESNEALHRMDEEMLYFIQKK
BglG            TQLMREMLQLIKFQFSLNYQEESLSYQRLVTHLKFLSWRILEHASINDSDESLQQAVKQN
ArbG    ♦       TRVMQEILNIVKYQLNLDYNEQAFSYHRFVTHLKFFAQRLLGRTPVFSEDESLHDVVKEK
BS reg          TKVMQEILSIVKYHFKIEFNEESLHYYRFVTHLKFFAQRLFNGTHMESQDDFLLDTVKEK
AbgG            TKKINDILNIVKYTYNLELDEKTLNYERFITHLRFFFKRLERKEITESEDNFLLEQVKNK
                *  . ...  .         * .  * * .***..   ..          *  .   ..

SacY            FKDAFLCALSIGTFVKKEYGFEFPEKELCYIAMHIQRFYQRSVAR----
SacT            YSFAYQCALELAEFLKNEYQLHLPESEAGYITLHVQRLQDLSE------
BglG            YPQAWQCAERIAIFIGLQYQRKISPAEIMFLAINIERVRKEH-------
ArbG            YTLAYHCAEKIQDHIMLHYDYTLTKEELMFLAIHIERVRSELQEQTAE-
BS reg          YHRAYECTKKIQTYIEREYEHKLTSDELLYLTIHIERVVKQA-------
AbgG            YTDAYECMLKIEKYLGQE----LSNEEQLYLMLHIQRVTTREK------
                .  *  *   .  .          *  .. ....*
```

Figure 4.11 Comparison of amino acid sequences of the *B. subtilis* SacY and SacT proteins, *E. coli* BglG, *E. chrysanthemi* ArbG and *C. longisporum* AbgG. The BSreg sequence is the *B. subtilis* regulatory protein involved in ß-glucanase regulation (see text). An asterisk below the sequence indicates an identical residue in all five sequences and a dot indicates a conservative change. Residues discussed in the text are in bold. CLUSTALW (97) was used to produce this alignment.

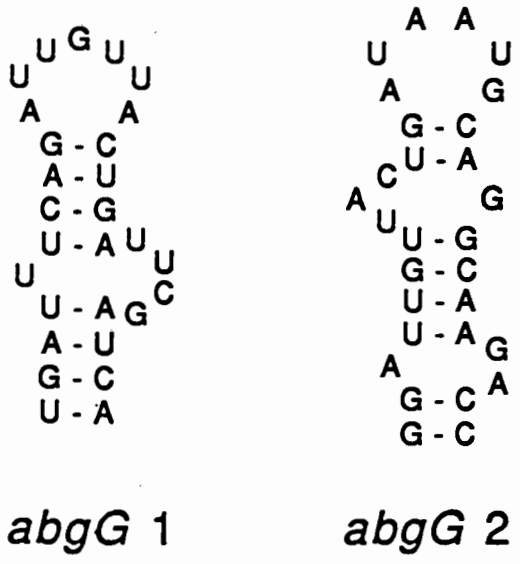


Figure 4.12 Proposed secondary structure formed by the Box A/ Box B sequence of *abgG*1 (ΔG -4.6) and *abgG*2 (ΔG -2.0). These structures were determined using the FOLD program.

Of interest is the homology of AbgG to a *B. subtilis* transcriptional antiterminator which regulates ß-glucanase synthesis (Table 4.2) (139,204). Whether the regulation of ß-glucanase synthesis in *C. longisporum* is also controlled by similar antiterminators remains to be determined.

abgF: Encoding a polypeptide of 616 amino acids (from position 3038 to 4885; 67.8 kDa), *abgF* is separated from *abgG* by 133 bp. The intergenic region contains the *abgG2* BoxA/ BoxB motif and terminator, discussed above. A putative RBS (GAGGAGAG) was located 7 bp upstream from the presumed translational start codon (AUG), and translation would terminate at the UAA stop codon (bp 4886-4889). No -10 or -35 regions were identified.

The predicted polypeptide shares sequence similarities with the phosphoenolpyruvate dependent phosphotransferase enzymes II of *E. coli* (*bglF*) (31,192), *E. chrysanthemi* (*arbF*) (85,86) and *Brevibacterium lactofermentum* (*ptsG*) (233) (Table 4.2, Figure 4.13). These transmembrane proteins mediate the transport of carbohydrates into the cytoplasm, as discussed in Chapter 1. In *E. coli*, BglF also regulates the antiterminator activities of BglG by phosphorylation, as discussed above (10,189).

Conserved features of these proteins include two phosphorylation sites, an amphipathic helical segment located at the N-terminus of the protein, and charged residues at the C-terminus (86,177). Putative AbgF phosphorylation sites were identified, around residues 26 and 538, based on sequence and location similarities to the BglF phosphorylation sites (Figure 4.13)(191). The first phosphorylation site of BglF, His-547 (corresponding in location to His-538 in AbgF), is essential for phosphorylation of enzyme IIA by HPr and is highly conserved in these two enzymes II and in ArbF (86). Site directed mutagenesis of BglF has also demonstrated that Cys-24 and His-306 of BglF (corresponding in location to Cys-26 and His-302 in AbgG) are involved in the transfer of the phosphoryl group to the transported sugar, with the Cys residue being phosphorylated (enzyme IIB) (86,191).

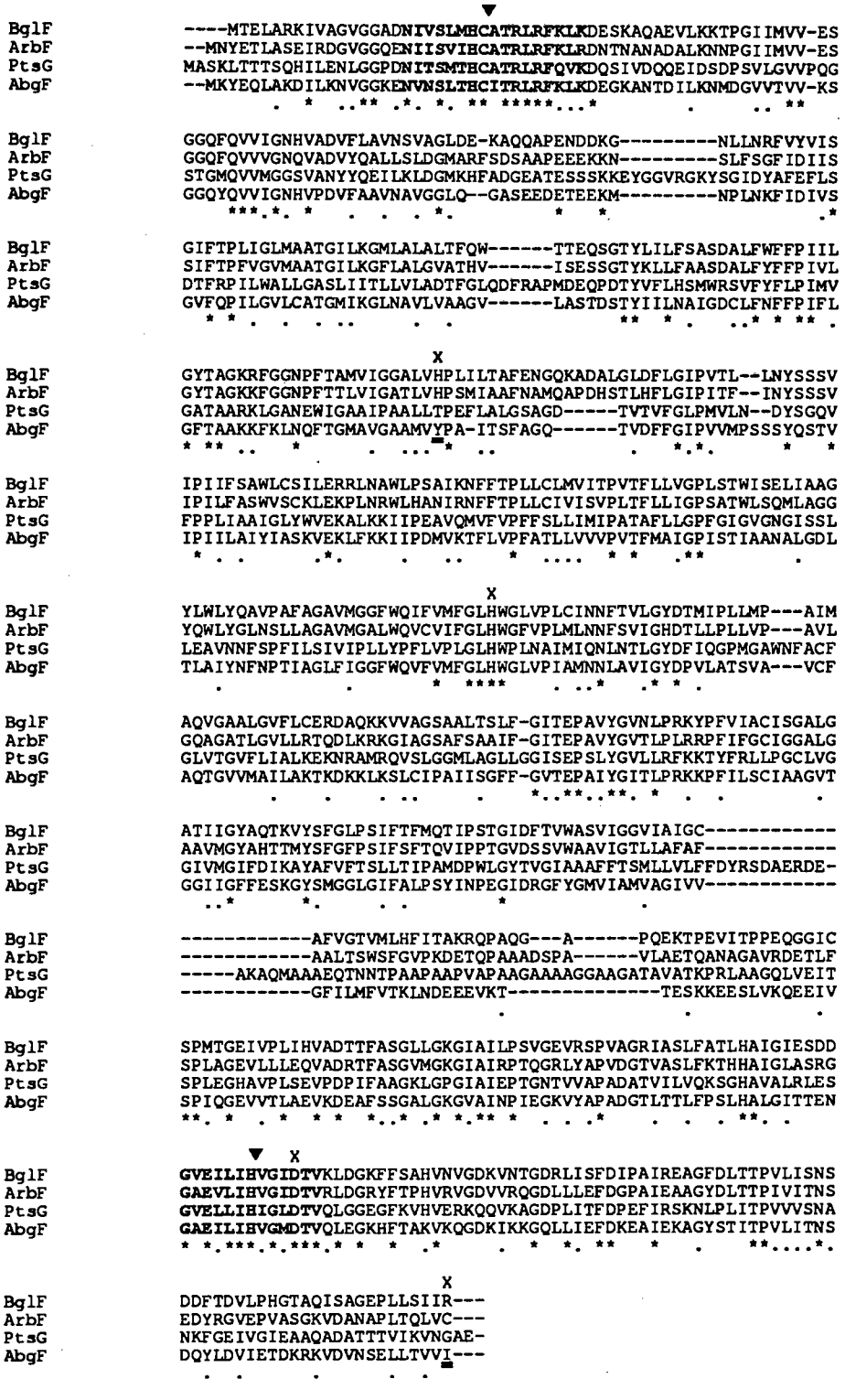


Figure 4.13 Amino acid alignment of enzyme II proteins from *E. coli* (BglF), *E. chrysanthemi* (ArbF), *B. lactofermentum* (PtsG) and *C. longisporum* (AbgF). Asterisks under the sequence indicate identical residues in all four sequences while the dots represent conservative changes. ▼ indicates phosphorylated residues, while X's indicate other residues discussed in the text. Phosphorylation sites are shown in bold. AbgF residues different to residues shown to be important in BglF function are double underlined. CLUSTALW (97) was used to create the alignment.

Although AbgF is very similar to BglF and ArbF, a number of differences were detected. His-183, shown to be important for substrate specificity of BglF and also conserved in ArbF is not present in AbgF (tyrosine at position 184) (86,191). Furthermore, only one of the other two residues in BglF, which were shown to function together with His-547 in phosphoryl transfer, were detected in AbgF (191). Asp-551 in BglF (which corresponded to Asp-542 in AbgF), but not Arg-625, was detected. Arg-625 was replaced with isoleucine (Ile-616) in AbgF and no other basic residues were located nearby on the C-terminus, which could have substituted for this function. This is similar to ArbF which also lacked the C-terminal Arg-625 (191). Assuming that AbgF functions in a similar fashion to BglF, and given the importance of these residues in BglF function, these differences are puzzling. Whether the residues present in AbgF could function in a similar manner to the BglF residues needs to be determined experimentally.

Analysis of the N-terminal domain of the predicted AbgF revealed the presence of a 14 amino acid helical amphipathic segment putatively terminated by a pair of glycyl residues (Figure 4.14). These amphipathic leader sequences are a characteristic of many PTS permeases and are thought to function in membrane insertion and bacterial envelope targeting (177). Similarly to ArbF, AbgF did not possess charged residues at the C-terminus (86).

As AbgF is a putative transmembrane protein an analysis of its overall hydropathy was undertaken (Figure 4.15). The strong hydrophobic nature of this polypeptide, characteristic of membrane proteins, and overall similarity to BglF, ArbF and PtsG suggests that AbgF is a transmembrane protein. Thus *abgF* appears to encode a transmembrane protein, similar to other PTS enzymes II, and is composed of three fused domains (ordered as enzyme domains II B, C and A) as discussed in Chapter 1.

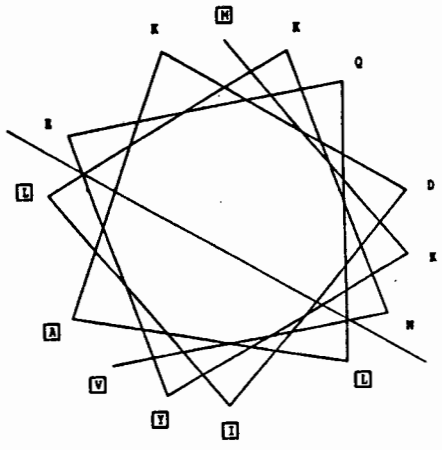


Figure 4.14 Helical wheel plot demonstrating the amphipathic nature of the first 14 N-terminal amino acids of AbgF. The line bisecting the plot separates hydrophobic (boxed) from hydrophilic residues. The plot was drawn using HELICALWHEEL of GCG (68).

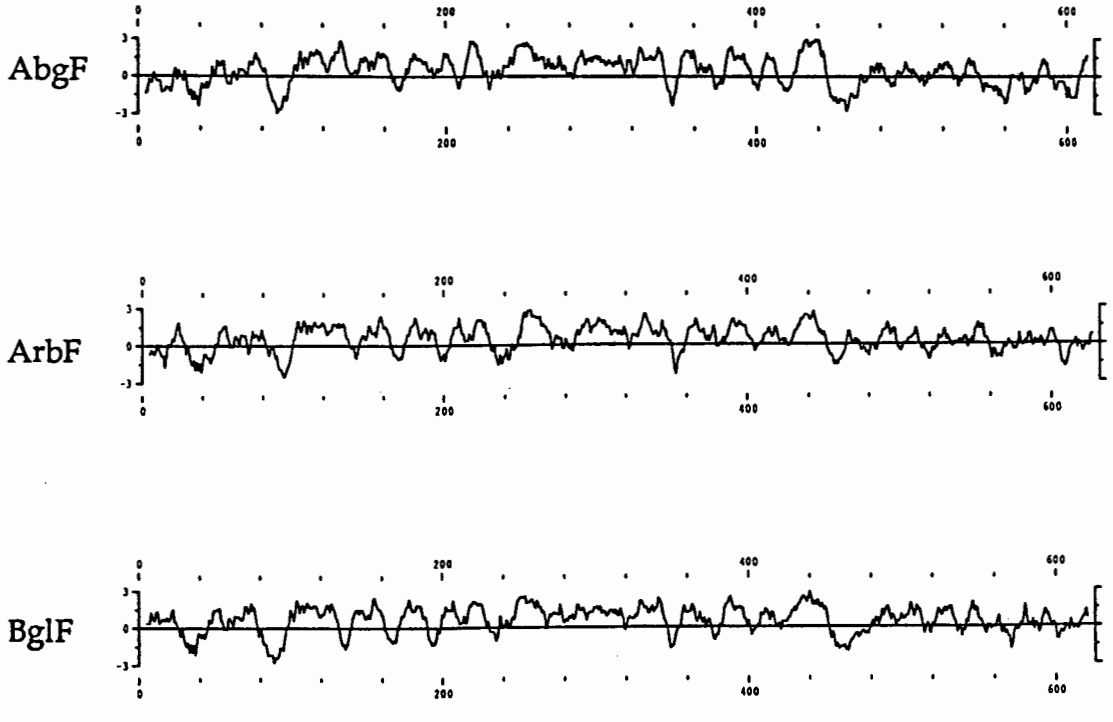


Figure 4.15 Hydropathy plot of the deduced AbgF amino acid sequence compared to ArbF and BglF. The curves are an average of residue specific hydrophobicity over a window of nine residues and were generated using PEPLOT (68) based on the method of Kyte and Doolittle (1982). Above zero is hydrophobic and below hydrophillic.

abgA: Separated from *abgF* by 30 bp, this ORF encodes a polypeptide of 473 amino acids (52 kDa) from bp 4916 to 6334. Seven bases upstream of the presumed AUG start codon (bp 4916 to 4919) is a putative ribosome binding site (AGGCGGA). No -10 or -35 regions could be identified. Translation presumably stops at the UAA stop codon (bp 6335 to 6338), and fourteen bp downstream from the UAA stop codon is a *rho*-independent transcription terminator (ΔG -17.3) (Figure 4.6).

Based on homology, *abgA* appears to encode a polypeptide similar to family I phospho-ß-glucosidases (EC 3.2.1.86) from *E. coli* (*ascB*, *bglB*) (81,192) and *E. chrysanthemi* (*arbB*) (85,86) (Figure 4.16). These enzymes are present in bacteria that accumulate saccharides as phosphates via the PTS, and cleave the phosphorylated sugar (88). For example, BglB, converts salicin-phosphate to glucose-6-phosphate and the phenol type aglycon (192) (see Figure 4.19).

Two conserved regions that have been detected in family I glycosidases (86,88,93) were also detected in AbgA (Figure 4.16). As discussed in Chapter 1, ß-glucosidases act by a general acid catalysis mechanism involving two acidic residues and recently the Glu-358 residue, invariant in all family 1 glycosidases, of an *Agrobacterium* glucosidase was shown to participate in glycosidic bond cleavage by acting as the nucleophile (93,225). This residue and a conserved sequence (LFIVENGLG, where E is the putative active site residue) was detected in AbgA. The other conserved region (FLWGGATAANQFEGA) of AbgA is located at the N-terminus, and is a signature sequence of family I glycosidases (18,86,93). Furthermore, all 27 invariant amino acids reported by Hassouni *et al.* (1992) were detected in AbgA. These results therefore suggest that *abgA* encodes a family I phospho-ß-glucosidase.

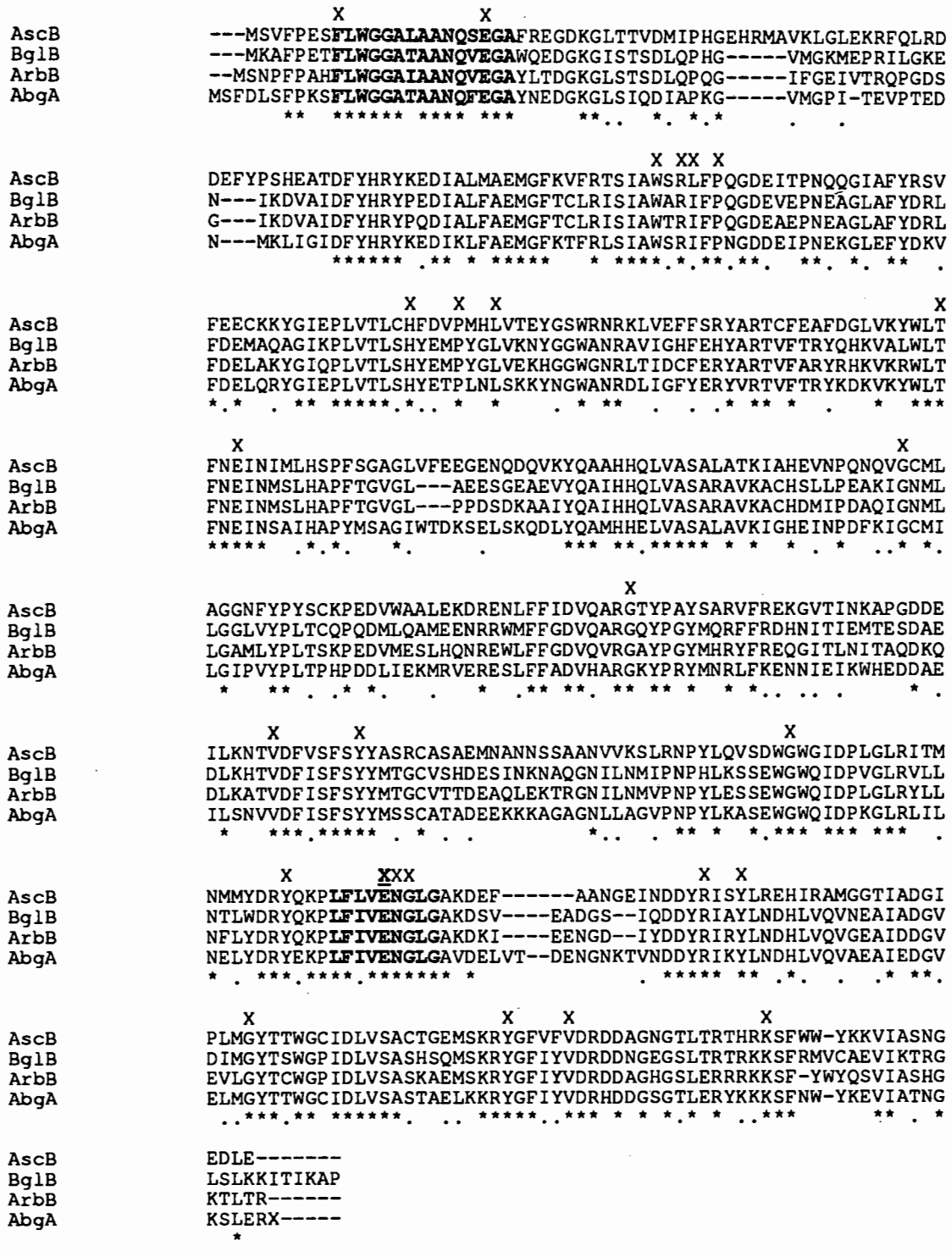


Figure 4.16 Alignment of AbgF with phospho-ß-glucosidases from *E. coli* (AscB, BglB) and *E. chrysanthemi* (ArbB). Asterisks below the sequence indicate identical residues in all sequences and dots indicate conservative changes. Conserved sequences mentioned in the text are in bold, while invariant residues described by Hassouni *et al.* (1992) are marked by an X. The underlined bold X indicates the active site residue discussed in the text. CLUSTALW (97) was used to produce this alignment.

ORF 6: This open reading frame is separated from *abgA* by 110 bp and potentially encodes a protein of 98 amino acids (from bp 6446 to 6739; 10.8 kDa) (Figure 4.6). Translation presumably starts from the AUG codon (bp 6446-6449) and ends at the UAA stop codon (bp 6740-6743). A putative ribosome binding site (AGGAGGG) 5 bp upstream from the AUG codon was identified. No transcription terminators were detected. One imperfect inverted repeat was identified internally in the gene which could possibly form a hairpin like structure (ΔG -7.6). No -10 or -35 regions were detected.

No homology to any other reported protein sequence was detected. The function of this putative protein in the cell, if any, remains to be determined.

glnB: This ORF, separated from ORF 6 by 118 bp, encodes a protein of 112 amino acids (from bp 6858 to 7193; 12.3 kDa). The gene is preceded by a potential ribosome binding site (AGGAAGUG), 8 bp upstream from the putative AUG start codon (bp 6858-6861). Translation could end at the UAA codon (bp 7194-7197) which is 17 bp upstream of a *rho*-independent terminator (ΔG -17.7). No -10 or -35 regions were detected (Figure 4.6)

The predicted protein of this gene shares sequence similarities with nitrogen P_{II} regulatory proteins (*glnB*) from *Synechococcus* sp. (strain PCC 7942) (212) and *Azospirillum braseilense* (235) and *B. subtilis* (229) (Figure 4.17). P_{II} proteins are involved in the regulation of enzymes involved in nitrogen metabolism (125) (Figure 4.18).

In enteric bacteria, P_{II} is uridylylated in nitrogen limiting conditions at a tyrosine residue by a uridylyltransferase (34,125) (Figure 4.18). The uridylylated P_{II} allows deadenylylation of glutamine synthetase (GS), which catalyses the conversion of glutamate and ammonia to glutamine (125). In nitrogen excess, P_{II} is deuridylylated and promotes adenylylation of GS and thereby controls the activity of this enzyme (34,125,197). P_{II} is also involved in the transcriptional control mediated by the *ntr* system, responsible for the reaction to nitrogen

availability (125,235). Phosphorylation of the regulator NtrC, in nitrogen limiting conditions, activates transcription of the *ntr* genes and is catalysed by NtrB. Dephosphorylation of NtrC by NtrB, and hence repression of the *ntr* genes, requires an interaction between NtrB and unmodified P_{II} which occurs in conditions of nitrogen excess (101,125,144,235).

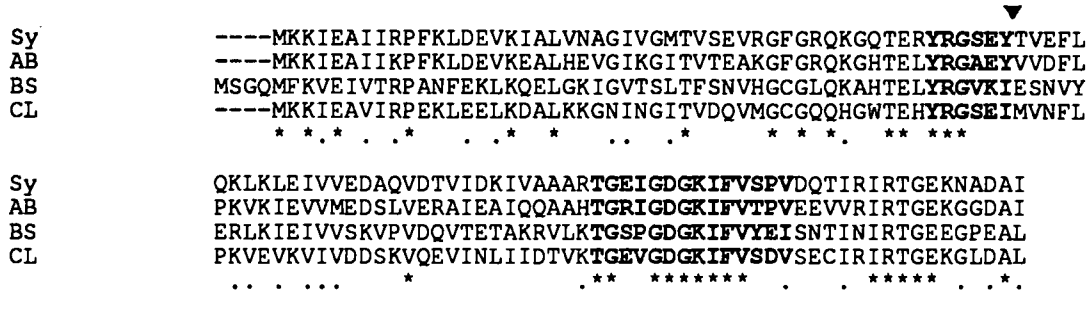


Figure 4.17 Sequence alignment of P_{II} proteins from *Synechococcus* (Sy), *A. braseilense* (AB), *B. subtilis* (BS) and *C. longisporum* (CL). Asterisks under the sequence indicate identical residues in all four sequences while dots indicate conservative changes. Sequences shown in bold are described in the text. The arrow indicates the site of uridylylation in enteric bacteria. CLUSTALW (97) was used to produce the alignment.

Two signature sequences were detected in the putative *glnB* gene product, P_{II} (Figure 4.17). The conserved C-terminal region of P_{II} enzymes (TGEVGDGKIFVSDV, residues 89-102) was detected (18), as well as a region described to be the uridylylation site in *E. coli* P_{II} (YRGSEI, from position 46 to 51). However, there is a significant difference between the consensus uridylylation site (Y-R-G-[AS]-E-Y) (18) and *C. longisporum* P_{II}, in that the second tyrosine residue is replaced with a isoleucine residue (position 51). This tyrosine residue, in *E. coli*, is the site of adenylylation and is conserved in the P_{II} enzymes of *Azospirillum* and *Synechococcus* (212,235).

During the course of these experiments another gene encoding a P_{II} protein (*nrgB*) was identified in *B. subtilis* (229). This gene and *C. longisporum glnB* are the first Gram positive P_{II} encoding genes isolated. It is interesting that although *C.*

longisporum P_{II} shares greater homology with the *Synechococcus* and *Azospirillum* proteins, both the predicted *C. longisporum* P_{II} and NrgB proteins had isoleucine residues at the uridylylation site (Figure 4.17). Whether this represents a conservative change in Gram positive bacteria, to retain some undetermined function is unknown. This change does, however, suggest that the *glnB* and *nrgB* gene products are not uridylylated at this site. It should also be noted that the P_{II} protein from *Synechococcus* possessed the conserved uridylylation site, yet was modified by phosphorylation of seryl residues (58), suggesting that other modification mechanisms are possible.

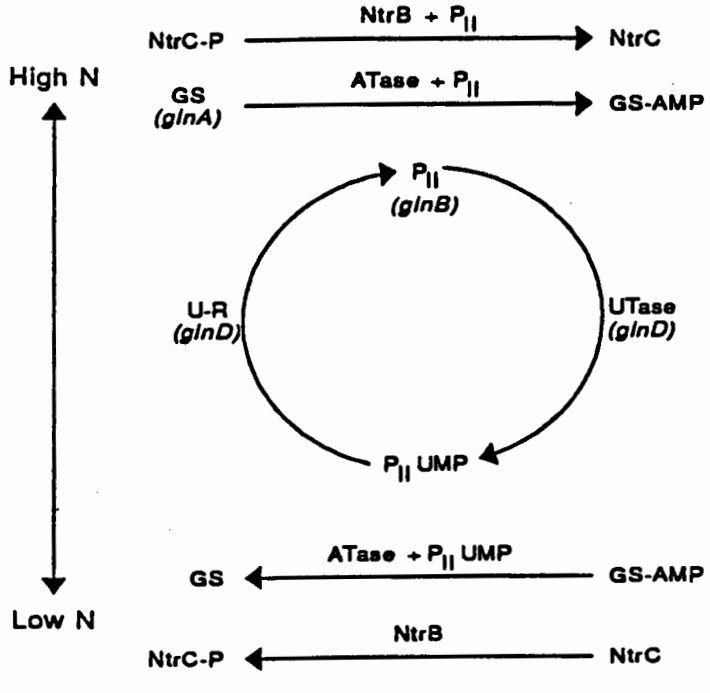


Figure 4.18 Schematic model for the regulating activities of enteric P_{II} enzymes in response to nitrogen (N). The uridylyltransferase (*glnD*) catalyses the uridylyation (UTase) and deuridylylation (U-R) of P_{II}. Adenylyltransferase (ATase) catalyses the adenylylation and deadenylylation of glutamine synthetase (GS). NtrB catalyses the phosphorylation and dephosphorylation of NtrC (from (101)).

No evidence for GS adenylation has been demonstrated in Gram positive bacteria (213,229). Although the *nrgB* gene product did appear to be involved in nitrogen metabolism, the function of the gene product remains unknown (229). The roles and functions of the *glnB* P_{II} protein should be ascertained in future research.

4.4.4 Organisation of the *abg* genes

Seven contiguous open reading frames were identified on genomic inserts cloned from *C. longisporum* and each ORF has been discussed above. Although interesting features, which require future investigation, were detected in almost all of the genes we were primarily concerned with the utilisation of carbohydrates involved in lignocellulose degradation. Thus the focus of the following research was directed towards the genes involved in the uptake and utilisation of aromatic-ß-glucosides.

The similarity in the organisation of the *abg* genes to those of the *Enterobaceriaceae* is striking, and suggests evolutionary conservation. A schematic diagram of the regulation of the *bgl* operon and the structural similarities to the *abg* system is shown in Figure 4.19. The organisation of the genes, including coding and non-coding regions, is so similar to those of *E. coli* that it is tempting to infer that the *abg* genes are organised in the form of an operon, and under the control of similar regulation mechanisms. However it is uncertain whether the *arb* genes of *E. chrysanthemi* function as an operon, primarily based on one major difference in the intercistronic region between *arbG* and *arbF* to those of the *bgl* operon genes (86,192).

A question could be asked as to how it was possible to detect MUCase activity in *E. coli* carrying pGDB2. This clone carries the terminator (*abgG* 2) but the antiterminator *abgG* is truncated. However, the activity obtained was most likely due to read though transcription (see Chapter 5). Given the dependency on vector promoters for expression (Chapter 3), it is unlikely that the gene is running off its own weak promoter, as is the case for BglF.

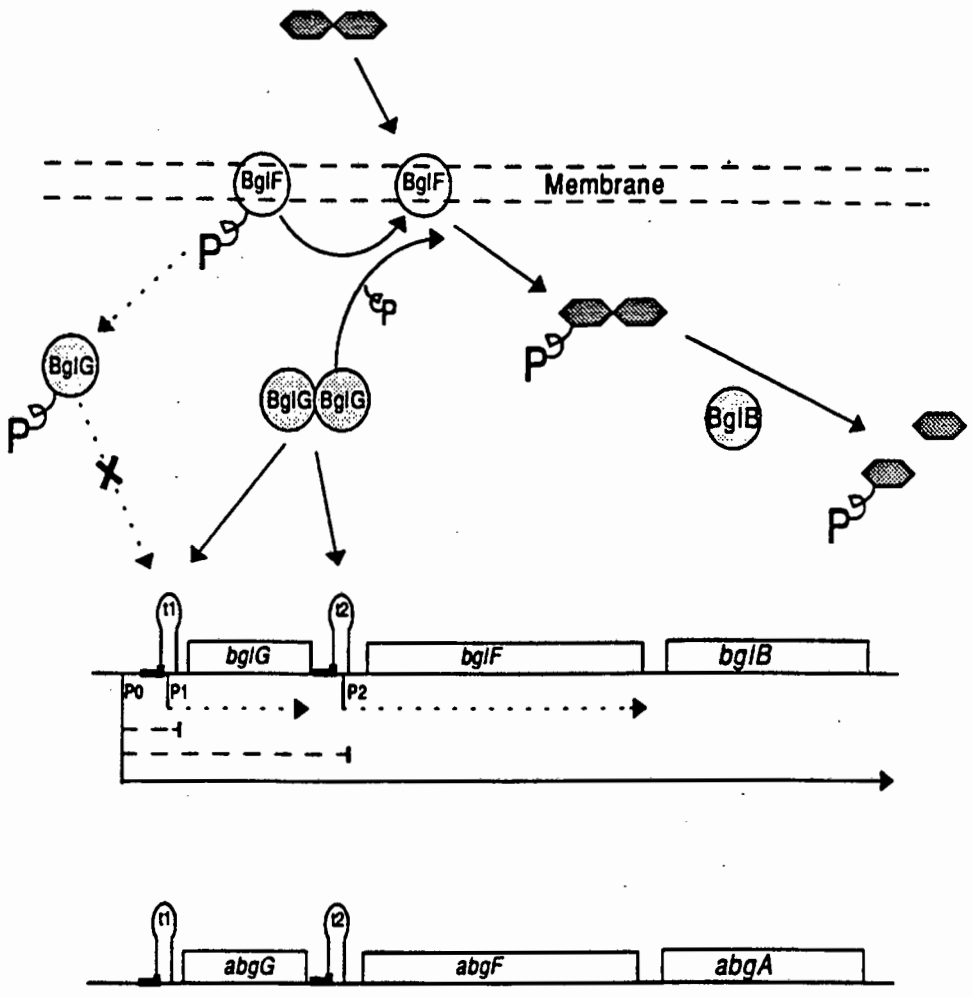


Figure 4.19 Schematic model of the regulation and genes of the *E. coli bgl* operon and a structural comparison to the *abg* gene system of *C. longisporum*. In the absence of substrate, transcription from P0 is terminated at *t1* and *t2*, in *E. coli*. Promoters P1 and P2 allow for basal levels of BglG and BglF. BglF acts as a negative regulator by phoshorylating BglG in the uninduced state and thereby inhibiting antiterminator activity. When the operon is induced by the substrate, BglF transports the carbohydrate into the cell and dephosphorylates BglG. This allows BglG dimerisation which then functions as an antiterminator, allowing transcription of the operon. BglB acts on the intracellular phospho-ß-glucosides. The position of the BoxA/ BoxB sequence is indicated by the thickened line preceding each terminator (from (12)).

Chapter 5

Characterisation of the aromatic-ß-glucoside uptake and utilisation genes from *C. longisporum*

5.1 Abstract

The function of each gene of the *abg* system and the involvement of the PTS system was examined here. Using mutants specific for PTS enzymes I and HPr, MUCase activity was shown to be dependent on the PTS system. Complementation of *E. coli* strains containing various mutations in the *bgl* operon and other genes, demonstrated that *abgF* and *abgA* did indeed encode a PTS dependent enzyme II and a phospho-ß-glucosidase, respectively. These genes were sufficient to confer on *E. coli* the ability to utilise arbutin and salicin for growth. Although the regulatory functions of the putative AbgG protein were not detected, the presence of a promoter 5' to *abgG* controlling the expression of the downstream gene, *abgF*, suggested that the *abg* gene system may be functioning as an operon. The ability of *C. longisporum* to utilise arbutin and salicin for growth was also demonstrated.

5.2 Introduction

The *E. coli* and *E. chrysanthemi bgl* and *arb* genes, respectively, comprise a system which can enable these organisms to utilise aryl-ß-glucosides for growth (85,192). The cryptic *E. coli bgl* operon is the best studied and, once activated, has been shown to be regulated by the constituent proteins, both in a positive and negative fashion (126), and by other factors (13). The transmembrane protein (BglF), which is dependent upon the PTS system for activity, negatively regulates another protein (BglG) by phosphorylation. BglG, acting as a dimer, is an antiterminator protein which prevents premature transcription termination and thereby acts as a positive regulator of the operon (10,12,102,127,188,189). In the presence of arbutin or salicin, BglF mediates the transmembrane transfer of these substrates into the cell with concomitant phosphorylation (191). This causes a dephosphorylation of BglG and allows its antiterminator actions (11). BglB, the third gene product of this operon, encoded by *bglB*, hydrolyses the intracellular phosphorylated substrate for subsequent metabolism (185). A schematic diagram of these activities was shown in Chapter 4 (Figure 4.19).

In previous chapters three *C. longisporum* genes, the *abg* genes, had been isolated, sequenced and appeared to encode an aryl-ß-glucoside uptake and utilisation system very similar to those of *E. coli* and *E. chrysanthemi*. These genes were the first such genes isolated from a Gram positive organism and, furthermore, were the first genes isolated which may link the use of phenolic glycosides to lignocellulose utilisation in the rumen. *C. longisporum* can hydrolyse lignocellulose as proficiently as other rumen bacteria (216,217) and has been shown to utilise at least one aromatic-ß-glucoside, salicin, for growth (215).

As the putative functions of each gene product had been deduced from sequence homology to the *bgl* and *arb* systems, these needed to be confirmed and characterised. These analyses are presented in this chapter. Furthermore, the ability of *C. longisporum* to utilise aromatic-ß-glucosides for growth was determined and the production of *abg* gene transcripts examined in this organism.

5.3 Materials and Methods

5.3.1 Bacterial strains and growth conditions

E. coli strains used for these experiments are listed in Table 5.1 and were cultured in 2X YT (182) at 30°C. Competent bacterial cells were prepared using the DMSO method of Chung and Miller (1988). *C. longisporum* B6405 (a gift from V.H. Varel (215)) was grown anaerobically in NRF medium (35) at 37°C with arbutin, salicin, glucose or cellobiose (0.1% w/v) as the major carbon source.

Table 5.1 *E. coli* strains used for these experiments

Strain	Genotype	Source
JM109	F⁻ *traD36 lacI*^{q} Δ(*lacZ*)*M15 proA*$^+$*B*$^+$/*e14*$^-$ (*McrA*$^+$) (Δ*lac-proAB*) *thi gyrA96* (*Nal*^{r}) *endA1 hsdR17* ($r_k^-m_k^+$) *relA1 supE44 recA1*	(230)
MK120	*arg met thi his xyl mtl rpsL bglA*$^-$ *bglB*$^-$ *bglR*$^+$ *bglF*^{c} *srlC::Tn10*	(151)
LP100	F⁻ *rpsL trp his argG met(A or B) ara leu lacY lacZ* Δ(*W4680*) *bgl-pho* Δ(201)*cel*Δ(100)	(151)
3300	*lacI22 relA1 spoT1 thi-1* λ⁻	(150)
CTR-7	*ptsI7 bglR10 relA1 spoT1 thi-1* λ⁻	(138)
DF-51	*lacI22 dctB3 ptsI2 relA1 spoT1 thi-1* λ⁻	(123)
1100	*bglR11 relA1 thi-1* λ⁻	(61)
1101	*his-62 ptsH1 bglR11 relA1 spoT1 thi-1* λ⁻	(61)

Growth analysis of the *E. coli* strains transformed with the various plasmid constructs (see below) was performed using M9 minimal medium (182) with relevant vitamins (0.005% w/v) and amino acids (0.04 mg/l) added. Carbon sources used include MUC, arbutin, salicin, glucose, cellobiose and fructose at final concentrations of 2 mg/ml. Growth, at 30°C, was analysed qualitatively on agar plates or quantitatively in liquid cultures, by measuring the absorbance at 600 nm. Transformed cells were cultured in the presence of the ampicillin (100 μg/ml) or chloramphenicol (50 μg/ml) or both as was necessary. To test for MUCase activity, MUC was added to 2X YT agar plates at a final concentration of 0.5 mM and activity was determined qualitatively by looking for fluorescence at 230 nm. IPTG (0.3 μM) was added to induce the *lac* promoter in *E. coli* strains overproducing *lacI*.

To determine the growth rates of *C. longisporum*, a mid- to late-log phase culture was diluted 1/1000 into fresh medium and total cell counts measured hourly. At late-log /early stationary phase serial dilutions were plated to compare the viable cell counts with the total cell counts.

5.3.2 Molecular methods and plasmids

All standard molecular methods were performed according to Sambrook *et al.* (1989). Restriction enzymes were obtained from Boehringer Mannheim and New England Biolabs and used according to the manufacturer's specifications. *In vitro* transcription and translation, using the *E. coli* S30 coupled transcription translation system from Promega Corporation, was performed according to the manufacturer's specifications.

Plasmids used in these experiments are listed in Table 5.2 and a schematic diagram of the constructs used is shown in Figure 5.1. pGDB2 and p8B1 (Chapter 3 & 4) were used to construct pBGL1 (Figure 5.2). Nested deletions used for sequencing (Chapter 4) were used in the construction of other plasmids, as shown in Figures 5.3, 5.4, 5.5 and 5.6.

Table 5.2 Plasmids used in these experiments.

Plasmid	Genes and Markers	Reference
pGDB2	Amp^r (Δ*abgG*) *abgF abgA* ORF6 *glnB*	Chapter 4
p4.9	Amp^r (Δ*abgG*) *abgF* (Δ*abgA*)	Chapter 4
p81	Amp^r *macA trsS abgG abgF* (Δ*abgA*)	Chapter 4
p8B1	Amp^r (Δ*macA*) *trsS abgG* (Δ*abgF*)	Chapter 4
pORF1.1 and p1C	Amp^r (Δ*abgG*) *abgF* (Δ*abgA*)	Results and Discussion
p18 and p19	Amp^r (Δ*abgF*) *abgA* (ΔORF6)	Results and Discussion
pORF1ACYC	Clm^r (Δ*abgG*) *abgF* (Δ*abgA*)	Results and Discussion
pBGL1	Amp^r (Δ*macA*) *trsA abgG abgF abgA* ORF6 *glnB*	Results and Discussion
p23A.1	Amp^r (Δ*trsA*) *abgG abgF abgA* ORF6 *glnB*	Results and Discussion
p445B.1	Amp^r *abgG abgF abgA* ORF6 *glnB*	Results and Discussion
p145D.1	Amp^r *abgG abgF abgA* ORF6 *glnB*	Results and Discussion
p7C.1	Amp^r (Δ*abgG*) *abgF abgA* ORF6 *glnB*	Results and Discussion
pUC18 / pUC19	Amp^r *lacZ*	(230)
pACYC184	Clm^r Tet^r	(36)

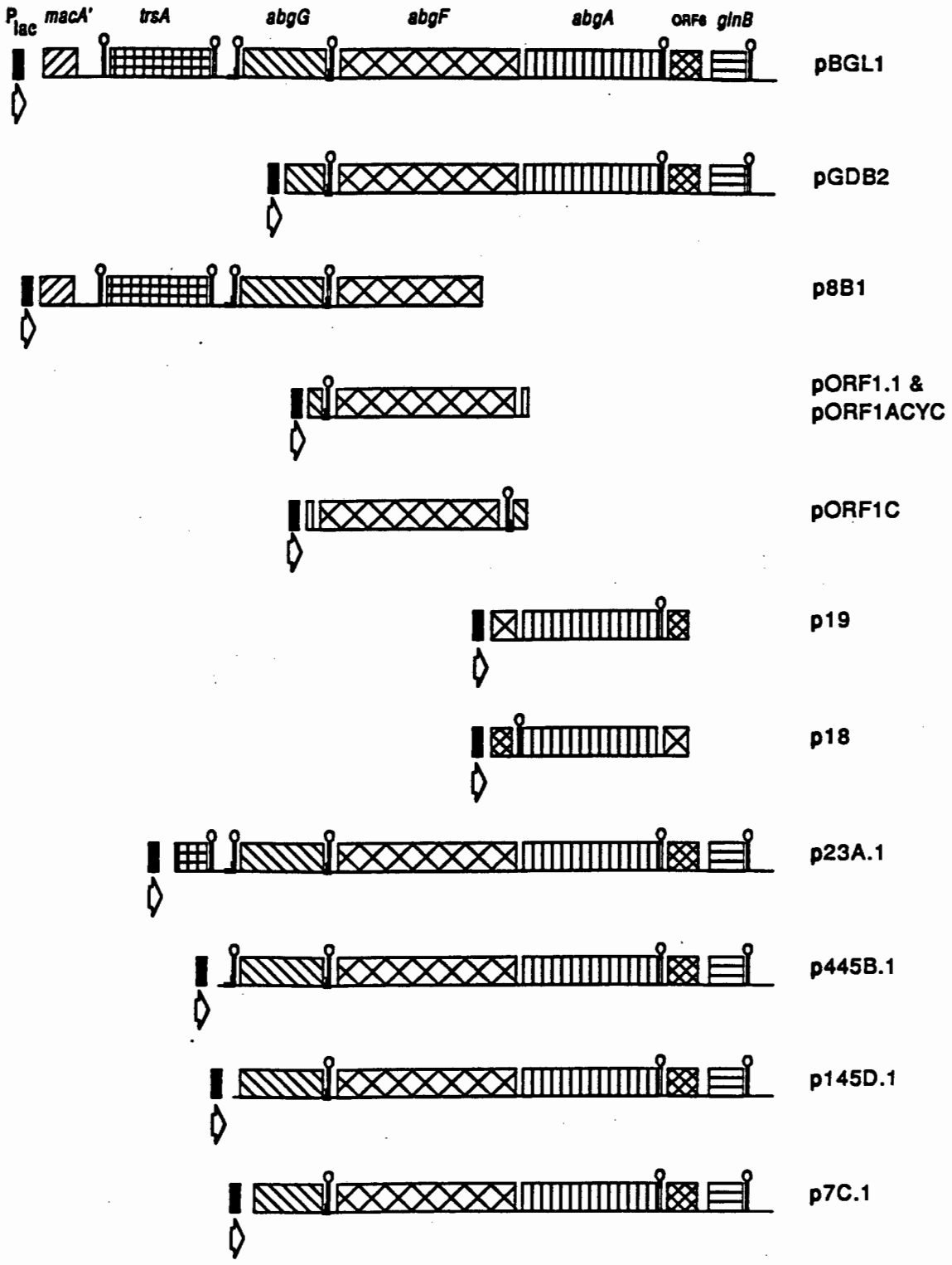


Figure 5.1 A schematic representation of the constructs used in these experiments. The position of the *lac* promoter is indicated.

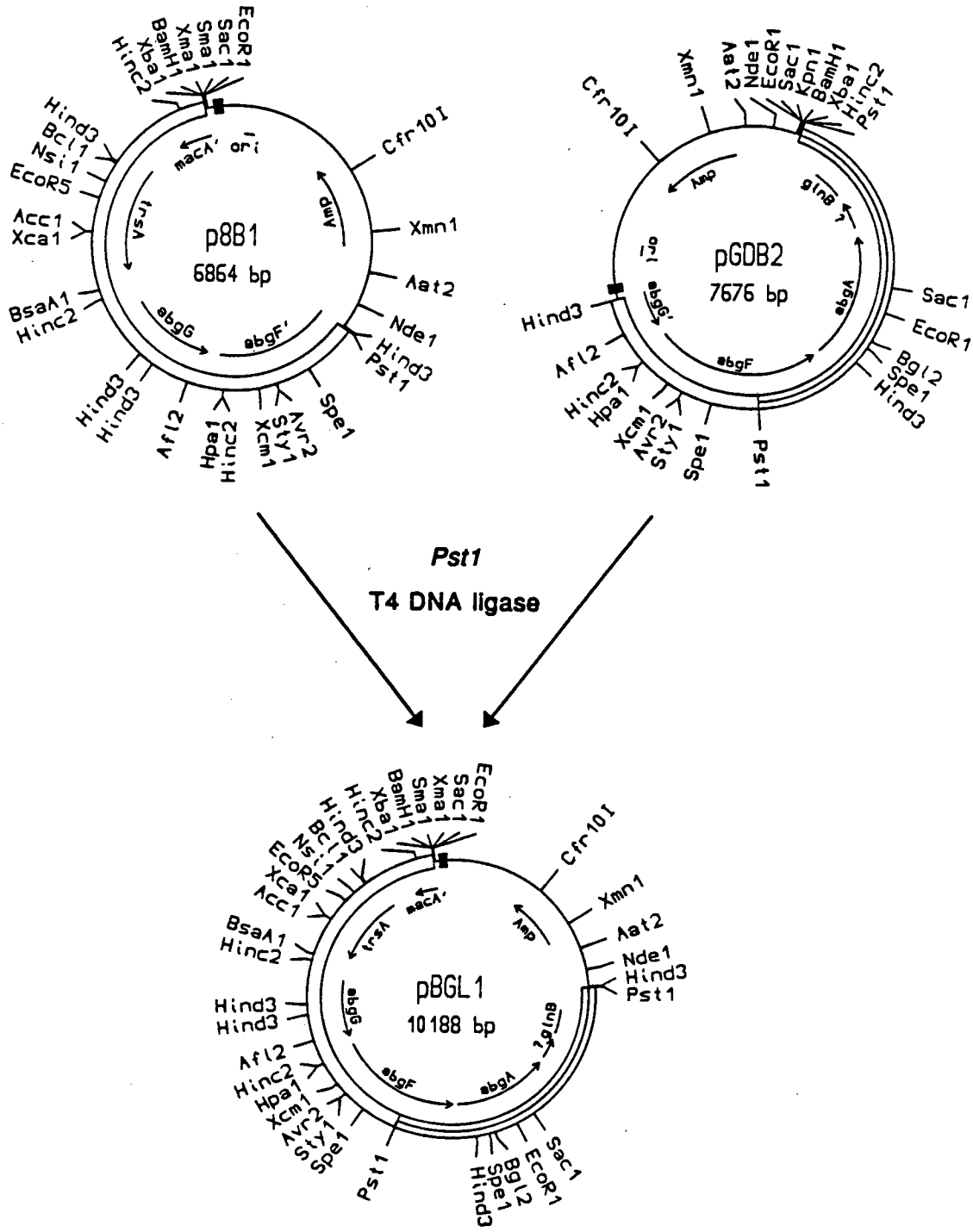


Figure 5.2 Construction of pBGL1 from p8B1 and pGDB2. The filled rectangle indicates the *lac* promoter.

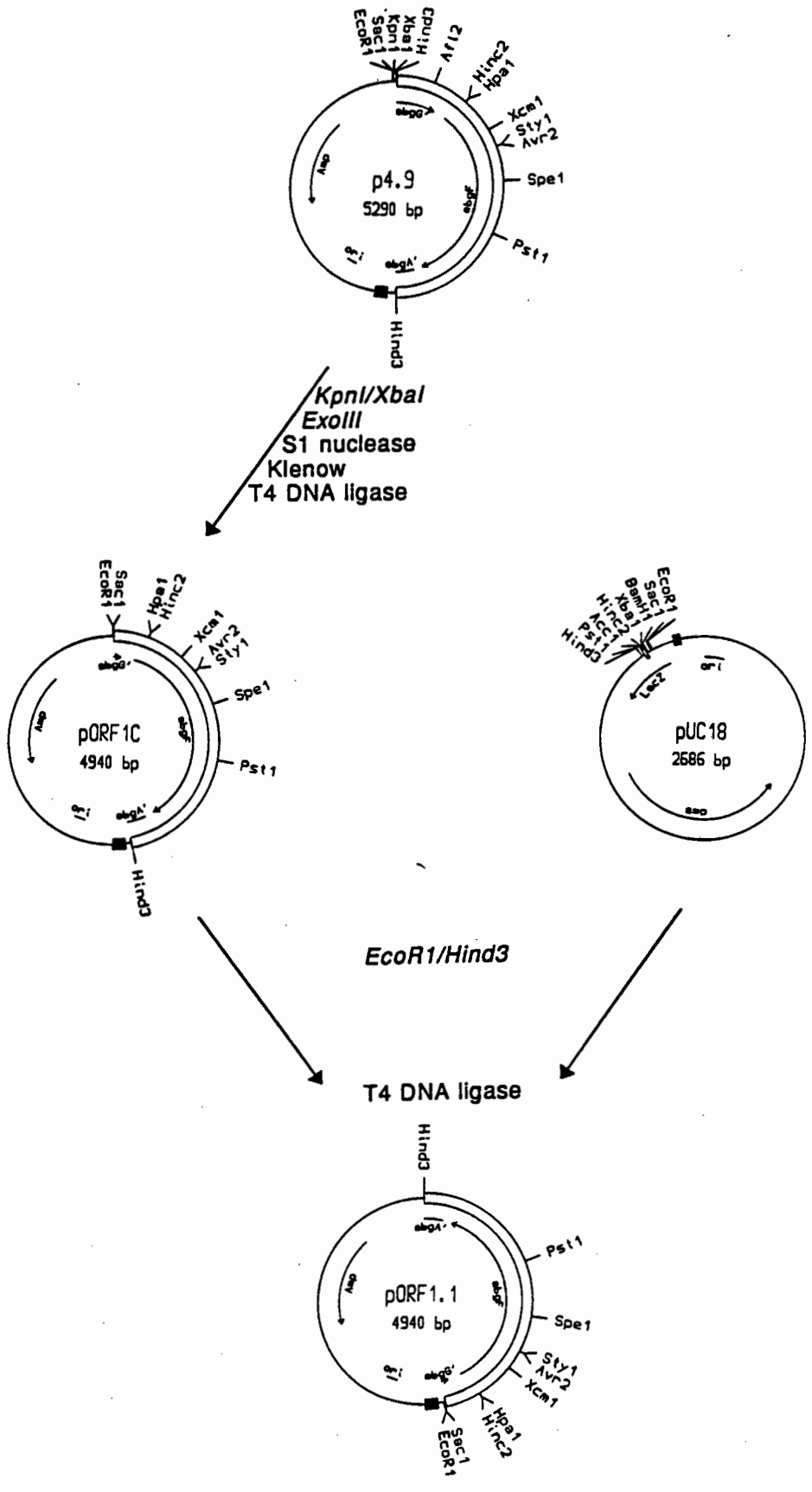


Figure 5.3 Construction of pORF1C and pORF1.1 containing the *abgF* gene. The filled rectangle indicates the *lac* promoter.

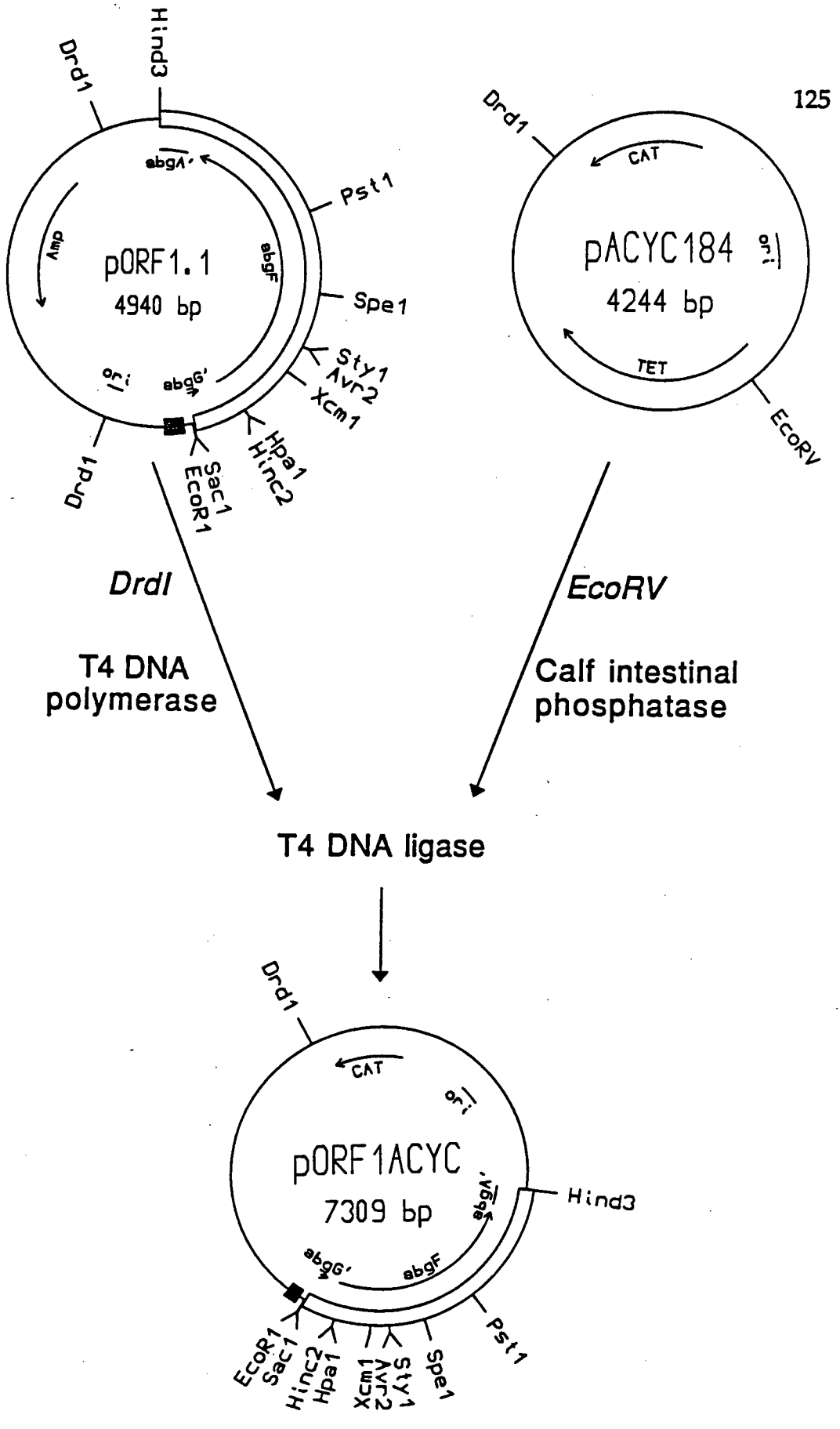


Figure 5.4 Construction of pORF1ACYC containing the *abgF* gene. The filled rectangle represents the *lac* promoter.

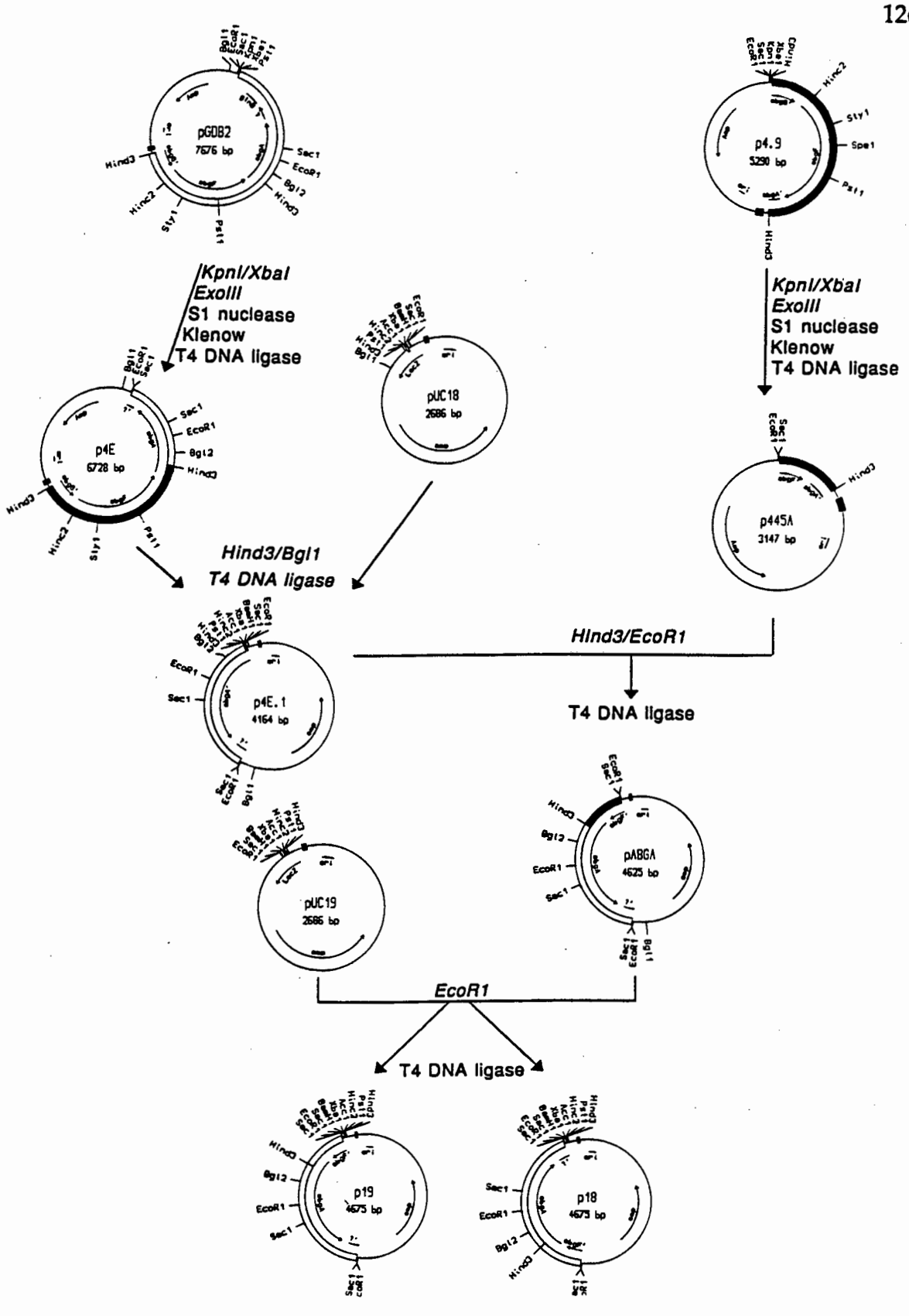


Figure 5.5 Construction of p18 and p19 containing the *abgA* gene. The filled rectangle indicates the *lac* promoter.

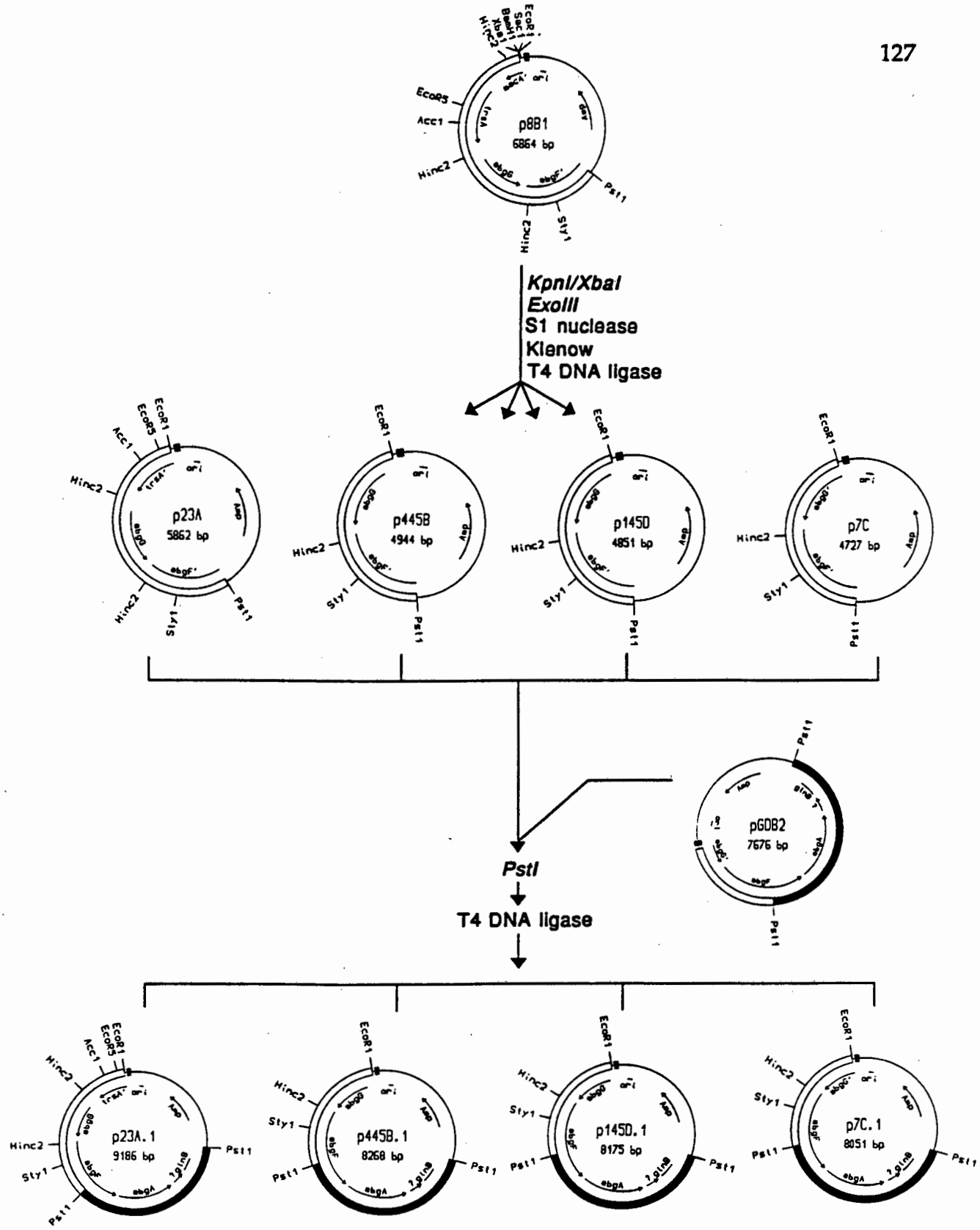


Figure 5.6 Construction of p23A.1, p445B.1, p145D.1 and p7C.1. The filled rectangle represents the *lac* promoter and the filled arc indicates the portion of pGDB2 contributed to the plasmid constructs.

C. longisporum RNA was extracted from cells in late-log phase according to the method of Aiba (1). RNA was treated with DNAse I and stored at -70°C. *E. coli* RNA was obtained from cultures grown in 2X YT medium using the same method. RNA was electrophoresed in formaldehyde 1.5% agarose gels with 1X MOPS running buffer (60). RNA, transferred to HybondN$^+$ membranes (Amersham, England) in 20X SSC (182), was probed with ^{32}P dATP or dCTP labelled DNA probes.

5.4 Results and Discussion

5.4.1 Involvement of the PTS system in MUCase activity

As the *abg* genes were similar to the PTS dependent *bgl* and *arb* systems, the initial analysis was to determine whether the *E. coli* PTS system was involved in the MUCase activity encoded on pGDB2. A number of mutant *E. coli* strains were used, including CTR-7 and DF-51, which contain different mutations in the *ptsI* gene, and 1101, which contains a mutation in the *ptsH* gene. These mutations block phosphoryl transfer from phosphoenolpyruvate to enzyme II. Thus if the PTS system were involved in MUCase activity, the activity should be abolished in these strains. Strains 3300 and 1100 are the parental strains, with wild type PTS genes, and were used as controls.

pGDB2 was transformed into these strains and grown on 2X YT medium containing MUC (Table 5.3). The loss of MUCase activity in DF-51 and CTR-7 strains clearly demonstrates the dependence on the PTS enzyme I for this activity. Although MUCase activity appeared to decrease in 1101 (the *ptsH* mutant), it was not completely abolished. It has been shown, however, that point mutations of HPr still retain significant HPr activity, although the growth on most PTS sugars is affected in strains possessing these mutations (176). Thus the presence of low levels of MUCase activity in this strain was probably a result of residual levels of HPr activity.

Table 5.3 Involvement of *E. coli* PTS enzyme I and HPr in the MUCase activity encoded on pGDB2. MUC activity was qualitatively measured. (+++ = strong activity; ++ = moderate activity; + = detectable activity; - = activity not detected)

Strain	pGDB2	pUC19
3300	+++	-
CTR-7	-	-
DF-51	-	-
1100	+++	-
1101	+	-

5.4.2 Substrate utilisation

To determine whether the genes encoded on pGDB2 could confer the ability to grow on soluble lignocellulosic substrates, *E. coli* LP100 in which both the cryptic *bgl* and *cel* operons have been deleted (185) was transformed with this plasmid. Transformants were streaked onto minimal media agar containing cellobiose, arbutin and salicin with fructose as a positive control (Table 5.4).

E. coli carrying pGDB2 was able to utilise salicin and arbutin, but not cellobiose, for growth. Thus the proteins encoded on pGDB2 are responsible for the uptake and utilisation of aromatic-ß-glucosides. Interestingly, *E. coli* (pGDB2) did not grow on medium containing MUC as the sole carbon source (results not shown). Although it is puzzling that pGDB2 encoded enzymes responsible for the MUCase activity yet was unable to utilise this substrate for growth, the observed activity could be explained if the enzyme(s) responsible were only cleaving the agluconic bond. This would cause the release of methylumbelliferone and cellobiose, neither of which could be utilised for growth. As discussed below, however, the *abgF* gene product may be solely responsible for this activity.

Table 5.4 Growth of *E. coli* LP100 transformed with pGDB2 or pUC19 on M9 minimal medium containing the carbon sources as indicated Growth was monitored after 72 hr. (+ = growth; - = growth not detected)

Plasmid	Fructose	Salicin	Arbutin	Cellobiose
pUC19	+	-	-	-
pGDB2	+	+	+	-

It should be noted that although other laboratory strains of *E. coli* contain cryptic *bgl* and *cel* operons, spontaneous activation to Arb^+ Sal^+ was occasionally observed. When this occurred, single colonies appeared, which grew well on the lignocellulosic substrates, irrespective of the plasmid present (results not shown). Thus growth was only recorded as positive if the entire streak demonstrated growth. Also, the production of a brown/black pigment in the medium was observed to occur whenever cells utilised arbutin as the sole carbon source (see Figure 5.8). This pigment was also produced by the spontaneous Arb^+ mutants. and results from the oxidation of hydroquinone, released during the hydrolysis of arbutin.

5.4.3 Characterisation of the *abg* gene products

To confirm the function of the gene products predicted for each open reading frame (Chapter 4), a number of constructs containing each of the *abg* genes were made (Figures 5.1 to 5.6). Furthermore, two *E. coli* strains were used which contain a number of mutations beneficial for this analysis (Table 5.1). *E. coli* LP100, mentioned earlier, has the entire *bgl* and *cel* operons deleted yet retains constitutive *bglA* phospho-ß-glucosidase activity specific for the hydrolysis of phospho-arbutin (185). *E. coli* MK120 has lost both *bglA* and *bglB* yet possesses a constitutive *bglF*, specific for the PTS dependent transport of arbutin and salicin (151).

abg F: To confirm that *abgF* encoded an enzyme II, pORF1.1 and 1C (carrying *abgF*) were transformed into *E. coli* LP100 and grown on medium containing the various substrates shown in Table 5.5. pORF1.1 was constructed such that transcription of the gene could run off the upstream *lac* promoter, while pORF1C was cloned in the reverse orientation respective to the promoter (Figure 5.1 and 5.3).

To grow on arbutin, LP100 requires a functional enzyme II to transport and phosphorylate the substrate. The phosphorylated substrate would then be hydrolysed by the constitutive *bglA* gene product and the products utilised further for growth. The ability of this strain, while carrying pORF1.1, to grow on arbutin indicates that the *abgF* gene does encode a functional enzyme II. Furthermore, transcription was dependent upon the vector promoter as pORF1C, which was cloned in the reverse orientation with respect to the *lac* promoter, was unable to confer the Arb^+ phenotype. In other *E. coli* strains occasional, but poor, growth was observed in cells carrying this plasmid (pORF1C), suggesting that some promotion of transcription was possible, which was independent of the *lac* promoter (results not shown). Although preceded by a Box A/ Box B sequence and a terminator (*abgG* 2), the fact that transcription of the gene was occurring suggested that the production of AbgF in this construct was dependent on readthrough transcription from the *lac* promoter.

To grow on arbutin or salicin, *E. coli* MK120 requires a functional phospho-ß-glucosidase, as it carries a functional permease. The inability of AbgF (on pORF1.1) to confer this ability to this strain, which lacks both *bglA* and *bglB* phospho-ß-glucosidases, demonstrates that, as expected, the transported aromatic glycoside was not being hydrolysed by AbgF (Table 5.5). Intriguingly, MK120 expressing *abgF* still appeared to be able to hydrolyse MUC, apparent by detectable fluorescence (Table 5.5). Although MUC activity is dependent upon the phosphoryl transfer of the PTS, shown previously, phosphorylation of MUC would not cause the agluconic group to fluoresce (Marc Claeyssens, personal communication). This ability of AbgF remains unresolved and could be examined

in future analyses. It may also explain the ability of pGDB2 to confer a MUCase$^+$ phenotype to *E. coli* while not allowing the organism to utilise the substrate for growth.

Table 5.5 Qualitative growth analysis of *E. coli* LP100 and MK120 on various carbon sources containing the plasmids as indicated. (+ = growth; - = growth not detected).

Strain	Plasmid	Fructose	Arbutin	Salicin	MUCase[a]
LP100	pUC18 or pUC19	+	-	-	-
	pGDB2	+	+	+	*
	pORF1.1	+	+	-	*
	pORF1C	+	-	-	-
MK120	pUC18 or pUC19	+	-	-	-
	pGDB2	+	+	+	*
	pORF1.1	+	-	-	*
	pORF1C	+	-	-	-

[a]Cells were grown on 2X YT agar plates containing 0.5 mM MUC. (* = activity; - = activity not detectable)

abgA: To confirm that *abgA* encoded a phospho-ß-glucosidase, which has activity on phospho-arbutin and phospho-salicin, p18 and p19 were constructed (Figures 5.1 and 5.5). These plasmids contain the *abgA* gene in both orientations downstream of the *lac* promoter and were examined for their ability to confer the Arb$^+$ Sal$^+$ phenotype to MK120 (Table 5.6).

MK120 possesses a constitutive *bglF* gene and will therefore be able to transport and phosphorylate salicin and arbutin. However, as the organism lacks both the *bglA* and *bglB* genes it will be unable to grow on these substrates unless a phospho-ß-glucosidase, specific for these substrates, is present in the cell. As shown in Table 5.6, the gene product of *abgA* was able to confer the Arb$^+$ Sal$^+$

phenotype to cells carrying this gene, indicating that *abgA* encodes a phospho-ß-glucosidase which is active on the phosphorylated forms of salicin and arbutin. It should be noted that although the other *E. coli* strains used in these experiments possessed a constitutive *bglA* gene (whose product is active on phosphorylated arbutin only), the ability to confer a Sal^+ phenotype to cells was taken as indicative of *abgA* expression.

Table 5.6 Qualitative growth analysis of *E. coli* MK120 on various carbon sources containing the plasmids as indicated.
(+ = growth; - = growth not detected).

Plasmid	Fructose	Arbutin	Salicin	MUCase[a]
pUC18 or pUC19	+	-	-	-
pGDB2	+	+	+	*
p18	+	+	+	-
p19	+	+	+	-
pORF1.1	+	-	-	*

[a]Cells were grown on 2X YT agar plates containing 0.5 mM MUC. (* = activity; - = activity not detectable)

The ability of both p18 and p19 to confer Arb^+ Sal^+ phenotypes indicates that transcription of *abgA* can run off its own promoter. No promoter regions were detected 5' to this gene in the initial sequence analysis (Chapter 4), and a reexamination of this region again did not reveal any potential *E. coli*-like -35 and -10 regions (Figure 4.6).

To further confirm the functions of *abgF* and *abgA*, *abgF* on pORF1ACYC (Figures 5.1 and 5.4) was supplied *in trans* to both p18 and p19 in LP100, which lacks both the *bgl* and *cel* genes (Table 5.7). The ability to grow on both arbutin and salicin provided further evidence that these two genes were involved in the uptake and utilisation of salicin and arbutin. These results also demonstrate that

abgA and *abgF* can function independently to complement the Arb$^+$ Sal$^+$ phenotype.

Table 5.7 Confirmation of *abgF* and *abgA* function in *E. coli* LP100. The strain, containing the various plasmids, was grown on minimal medium with the various carbon sources as indicated. (+ = growth, - = growth not detected).

Plasmids	Fructose	Arbutin	Salicin
pUC19	+	-	-
pGDB2	+	+	+
pORF1ACYC/ p18	+	+	+
pORF1ACYC/ p19	+	+	+
pORF1ACYC/ pUC18	+	+	-
p18	+	-	-
p19	+	-	-

abgG: To determine whether *abgG* encodes a functional regulatory protein, a plasmid containing all the genes of the *abg* system was constructed (Figures 5.1 and 5.2). The resultant plasmid, pBGL1, contained the entire 7.5 kb fragment isolated on pGDB2 and p8B1 (Chapter 4) and thus contained all of the *abg* genes as well as the upstream and downstream genes.

The ability of this construct to confer the Arb$^+$ Sal$^+$ phenotype to *E. coli* was examined in JM109 (Table 5.8). This *recA*$^-$ strain was used as rearrangements of p8B1 were shown previously to occur in JM105 carrying p8B1 (Chapter 4). Furthermore, as JM109 overproduces *lacI*, the *lac* promoter could be controlled using IPTG. In comparison to pGDB2, JM109 carrying pBGL1 was able to grow on both arbutin and salicin without IPTG induction. As *abgF*, on pGDB2, was dependent on the vector promoter for expression these results indicated that a region(s) 5' to *abgF* on pBGL1, was involved in promotion of transcription. It

should be noted that weak growth of *E. coli* carrying IPTG inducible *abgF* (dependent on the *lac* promoter for expression) on aryl-ß-glucosides was observed without added IPTG, but only after prolonged incubation.

Table 5.8 Growth comparison of JM109 carrying pBGL1, pGDB2 or pUC18 on aryl-ß-glucosides with glucose as a control. Growth was measured after 48 hr. (+ = growth; - = growth not detected)

IPTG induction	Plasmid	Arbutin	Salicin	Glucose
YES	pBGL1	+	+	+
	pGDB2	+	+	+
	pUC18	-	-	+
NO	pBGL1	+	+	+
	pGDB2	-	-	+
	pUC18	-	-	+

Although it is tempting to interpret these data such that both the *abgG* gene and its 5' promoter are required for IPTG-independent transcription of the *abg* genes, further experiments were performed to test this hypothesis. Constructs were made from the shortenings used to sequence p8B1 which possessed various deletions of the region 5' and into *abgG* but contained all of the downstream genes (Figures 5.1 and 5.6). The growth of JM109, transformed with these constructs, on arbutin and salicin minimal media was examined with and without IPTG induction (Table 5.9, Figures 5.7 and 5.8). Construct p23A.1, with the 5' region and half of *trsA* deleted, was still able to confer growth on arbutin and salicin without added IPTG. This indicated that *trsA* and upstream regions did not contain the promoter regions responsible for these growth characteristics. In contrast, p445B.1, containing a deletion of all the 5' genes and the *trsA* terminator up to 32 bp before the *abgG*1 BoxA/ BoxB sequence, was unable to confer the Arb^+ Sal^+ phenotype without added IPTG. Similar results were obtained with the other deletions, both 3' to the *abgG*1 Box A / Box B sequence (p145D.1) and a

deletion into *abgG* itself (p7C.1). This suggested that transcription in these constructs was dependent on the *lac* promoter.

Table 5.9 Growth comparison of JM109 carrying various *abg* constructs, with and without IPTG. Growth was monitored after 72 hr. (+ = growth; - = no growth).

IPTG Induction	Plasmid	Arbutin	Salicin	Glucose
NO	pBGL1	+	+	+
	pGDB2	-	-	+
	p23A.1	+	+	+
	p445B.1	-	-	+
	p145D.1	-	-	+
	p7C.1	-	-	+
	pUC18	-	-	+
YES	pBGL1	+	+	+
	pGDB2	+	+	+
	p445B.1	+	+	+
	p145D.1	+	+	+
	p7C.1	+	+	+
	pUC18	-	-	+

These results also suggested that the *abg* promoter region is located 5' in the upstream sequence deleted in p445B.1, but 3' to the upstream sequence deleted in p23A.1. If the *abg* genes are regulated in a similar fashion to the *bgl* operon of *E. coli*, then the promoter of these genes is located 5' to the *abgG*1 BoxA/ BoxB sequence. Although the results presented here do indicate that such a promoter is present, this region was not identified in the AT-rich sequence

(Chapter 4) and a reexamination did not reveal any *E. coli* promoter like regions. The presence of a 5' promoter region which controls downstream genes does, however, suggest that the *abg* gene system is functioning as an operon.

Although the presence of a promoter 5' to *abgG* has been shown, the regulatory functions of the protein encoded by this gene have not been detected. Furthermore, as expression of *abgF* and *abgA* occurred in *E. coli*, carrying pGDB2, the expression of these genes did not appear to be dependent on the proposed regulatory activities of *abgG* (Chapter 4). To determine whether AbgG was influencing the *abg* genes in *E. coli*, the growth rates of JM109 carrying various *abg* constructs were analysed (Figure 5.9). These constructs all contained the *abgF* and *abgA* genes but possessed various deletions of *abgG* and upstream sequences, as described earlier. Salicin was chosen as the carbon source to indicate AbgA activity, IPTG was present to induce the *lac* promoter and control cultures were grown without IPTG.

In the presence of IPTG, there was no observable differences in the growth rates of *E. coli* carrying any of constructs tested. Although the growth rates of cells carrying pBGL1 and p23A.1 appeared to be adversely affected when grown in the presence of IPTG, the growth rates of these cultures, grown without added IPTG, were more comparable. Thus the lack of a significant difference in the growth rates of cells carrying these constructs, specifically pGDB2 and p7C.1 which contain deletions of the *abgG* gene itself, suggests that the *abgG* gene product was not regulating the *abg* genes in *E. coli*.

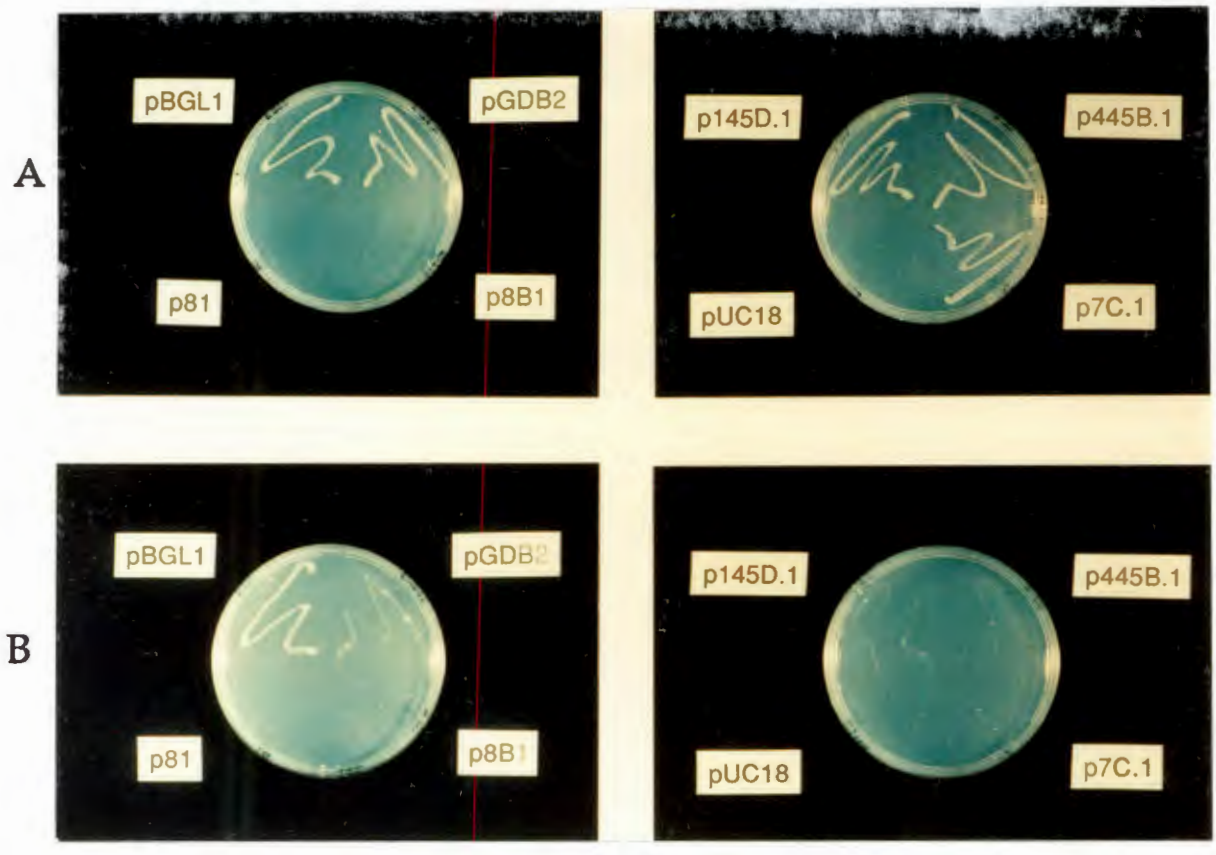


Figure 5.7 Comparison of *E. coli* JM109, carrying the various plasmids as indicated, growing on salicin with (A) and without (B) IPTG.

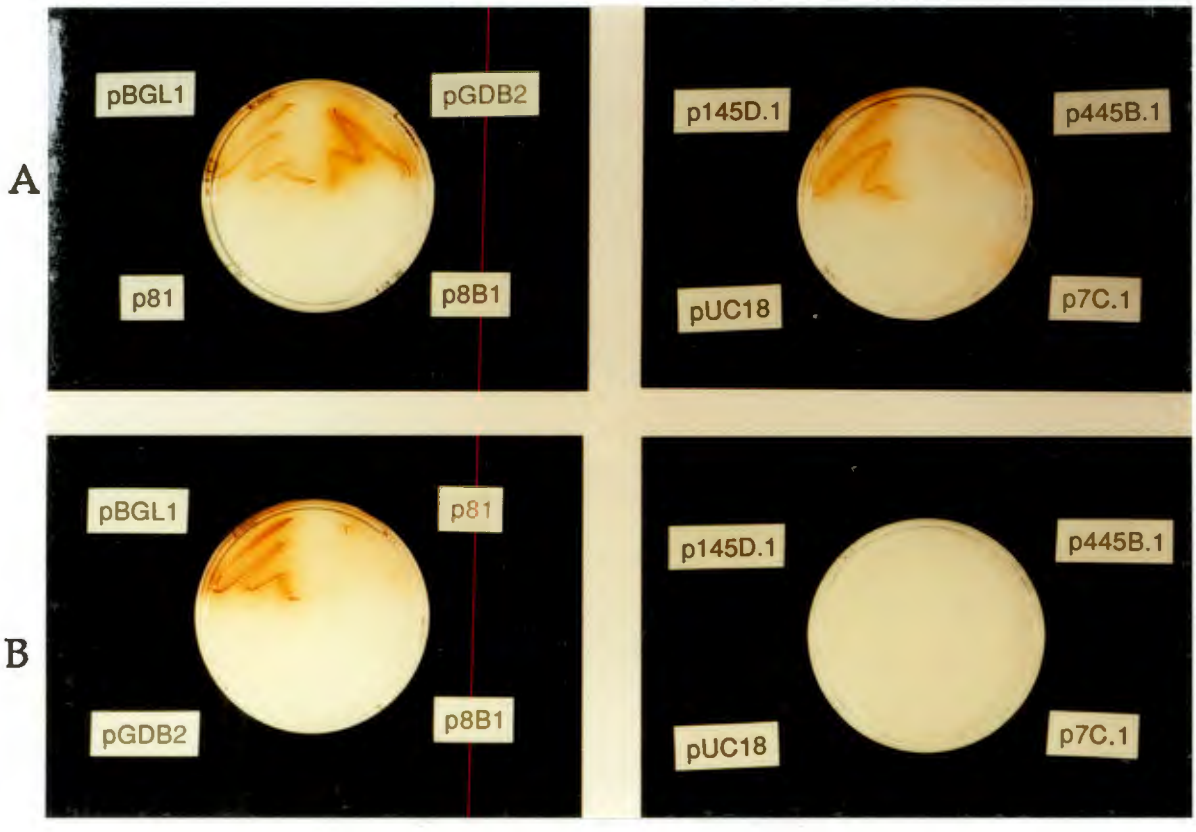
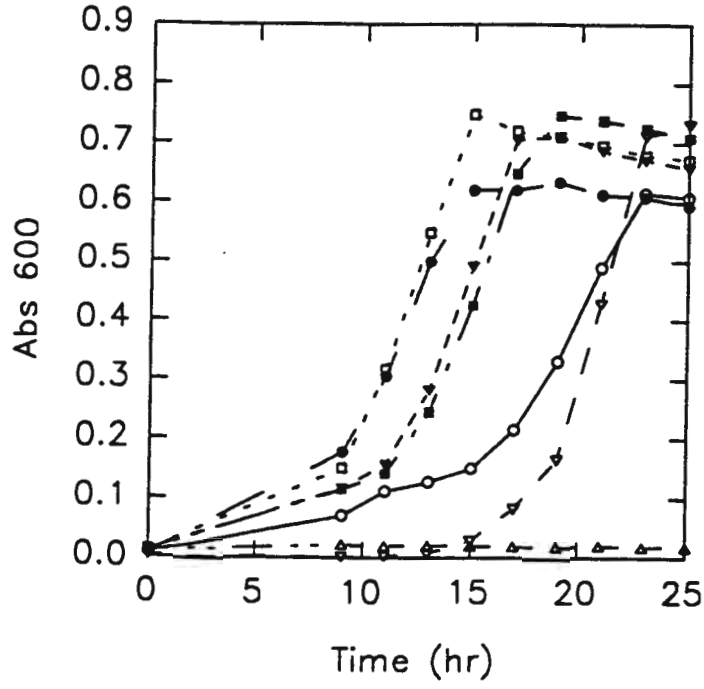


Figure 5.8 Comparison of *E. coli* JM109, carrying the various plasmids as indicated, growing on arbutin with (A) and without (B) IPTG. The oxidized hydroquinone is clearly visible as a brown pigment.

A



B

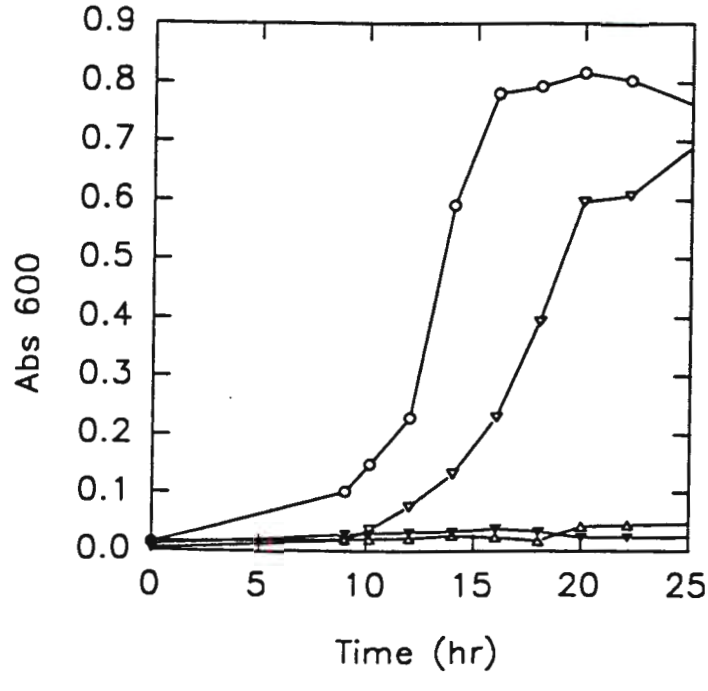


Figure 5.9 Growth analysis of *E. coli* JM109 carrying various *abg* constructs on salicin minimal medium with (A) and without (B) IPTG. Symbols: Δ pUC18; o, pBGL1; ●, pGDB2; ▽, p23A.1; ▼, p445B.1; □, p145D.1; ■, p7C.1.

There are a number of questions raised by these results which need to be addressed in future analyses. As *abgF* (on pGDB2) can be expressed, even though possessing a 5' terminator region (the *abgG* 2 sequence), the influence of these terminator structures on transcription in *E. coli* needs to be examined further. Also, as translation of the *abgG* gene transcript is dependent upon an unusual start codon (Chapter 4), and as the expression of this gene was not demonstrated in *E. coli*, the production of a functional protein product needs to be examined.

5.4.4 Confirmation of PTS involvement in the Arb$^+$ Sal$^+$ phenotype

MUC was not utilised as a substrate for growth by *E. coli* but was hydrolysed in a PTS dependent fashion, as discussed earlier (section 5.4.1). Although not able to hydrolyse salicin or arbutin, AbgF appeared to be at least partially responsible for the hydrolysis of MUC, in a phospho-ß-glucosidase independent fashion (section 5.4.3). As these results appeared contradictory, the involvement of the PTS system was reexamined (Table 5.10). Using the PTS mutants described before, which carry pGDB2, the involvement of the PTS should be demonstrated by an inability to utilise arbutin and salicin. As the *ptsI* mutants CTR-7 and DF-51, transformed with pGDB2, were unable to grow on arbutin or salicin, the involvement of the PTS system in the Arb$^+$ Sal$^+$ phenotype was confirmed.

Table 5.10 Confirmation of the involvement of the PTS system in the Arb$^+$ Sal$^+$ phenotype. Growth was monitored after 96 hr. (+ = growth, - = growth not detected)

Strain	Plasmid	Arbutin	Salicin	Fructose
3300	pGDB2	+	+	+
	pUC18	-	-	+
DF-51	pGDB2	-	-	+
	pUC18	-	-	+
CTR-7	pGDB2	-	-	+
	pUC18	-	-	+

5.4.5 *In vitro* transcription and translation

To confirm the sizes of the proteins predicted from the various *abg* genes sequences, *in vitro* transcription and translation was performed using the various constructs described previously (Figure 5.10). TrsA and AbgA proteins were detected and were of similar molecular weights to the proteins predicted from the sequence, 37 and 52 kDa respectively (Chapter 4). In all the constructs containing the *abgA* gene, the synthesis of smaller molecular weight proteins were also detected, but as a protein of the predicted molecular weight was present these smaller proteins are probably artifactual.

Proteins of the predicted molecular weight were not produced with constructs containing either the *abgG* or *abgF* genes. The inability to obtain AbgG (31 kDa) or AbgF (68 kDa) proteins *in vitro* may be due to transcription termination caused by the upstream terminators. Smaller molecular weight proteins were, however, detected in constructs containing *abgF*. The significance of these proteins remains to be determined but, as for the smaller proteins observed with AbgA, these may also be artifactual. Although *in vitro* transcription translation analysis was also performed with pBGL1, AbgF and AbgG were not detected (results not shown).

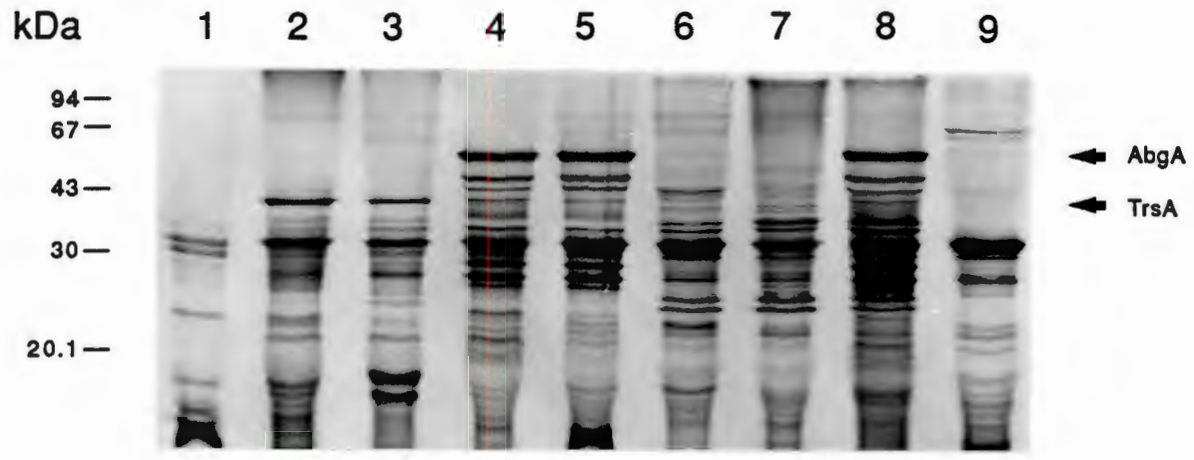


Figure 5.10 *In vitro* transcription and translation of constructs containing the *abg* genes. ^{35}S-methionine labelled *de novo* proteins were run on a 12.5% PAGE gel, which was dried and autoradiographed. The autoradiograph was overexposed (24 hr) in an attempt to identify AbgG and AbgF. Arrows indicate proteins of interest. Lanes 1, pECOR251; lane 2, p8B1; lane 3, p81; lane 4, p19; lane 5, p18; lane 6, pORF1C; lane 7, pORF1.1; lane 8, pGDB2; lane 9, pUC19.

5.4.6 Characteristics of *C. longisporum*

The isolation of an aryl-ß-glucosidase operon from *C. longisporum* and the link to phenolic glycoside utilisation in the rumen, prompted a brief analysis of the growth abilities of this organism on these substrates (Figure 5.11). *C. longisporum* is, however, notoriously difficult to maintain and grow and was once lost from the culture collection (215). Although we were able to routinely grow this organism some difficulties were encountered including its sensitivity to the anaerobic gas mixture used (results not shown) and pH (see below). In contrast to the results reported by Varel (1989), spore formation was not observed at any time during these experiments.

Another complicating factor was different morphologies that were obtained when routinely culturing this organism (results not shown). Although easily identifiable through the production of a red/orange extracellular polysaccharide, the organism displayed two different colony morphologies, described by the formation of either discrete or diffuse colonies on agar plates. These morphologies also had different growth characteristics on different substrates (results not shown). In liquid culture with cellobiose as the major carbon source, cells of the discrete type tended to aggregate and produce more extracellular polysaccharide while the other morphological type grew better and cells were diffused throughout the medium. As the aim of this analysis was to characterise the growth of *C. longisporum* and to establish growth conditions for RNA isolations, the diffuse morphology was chosen for further analysis. This choice was based on the organism's ability to grow well in culture and that its diffuse nature would allow for easier characterisation. The reason(s) for the occurrence of these different morphologies was not determined.

The diffuse type *C. longisporum* was able to grow on arbutin, salicin, glucose and cellobiose (Figure 5.11). Although growth on ß-glucan and cellulose was obtained it was substantially slower than on the other substrates tested (results not shown). During growth on the former substrates, cells had periods of motility which occurred mainly during mid-log phase. The size of the cells also appeared to vary, from short to very long rods (not shown). The viable cell counts, measured only at late-log phase, are shown in Table 5.11. Once the pH reached 5.5 the cells underwent rapid autolysis and, after total cell lysis, the culture medium became very viscous and retained the pigmented extracellular polysaccharide.

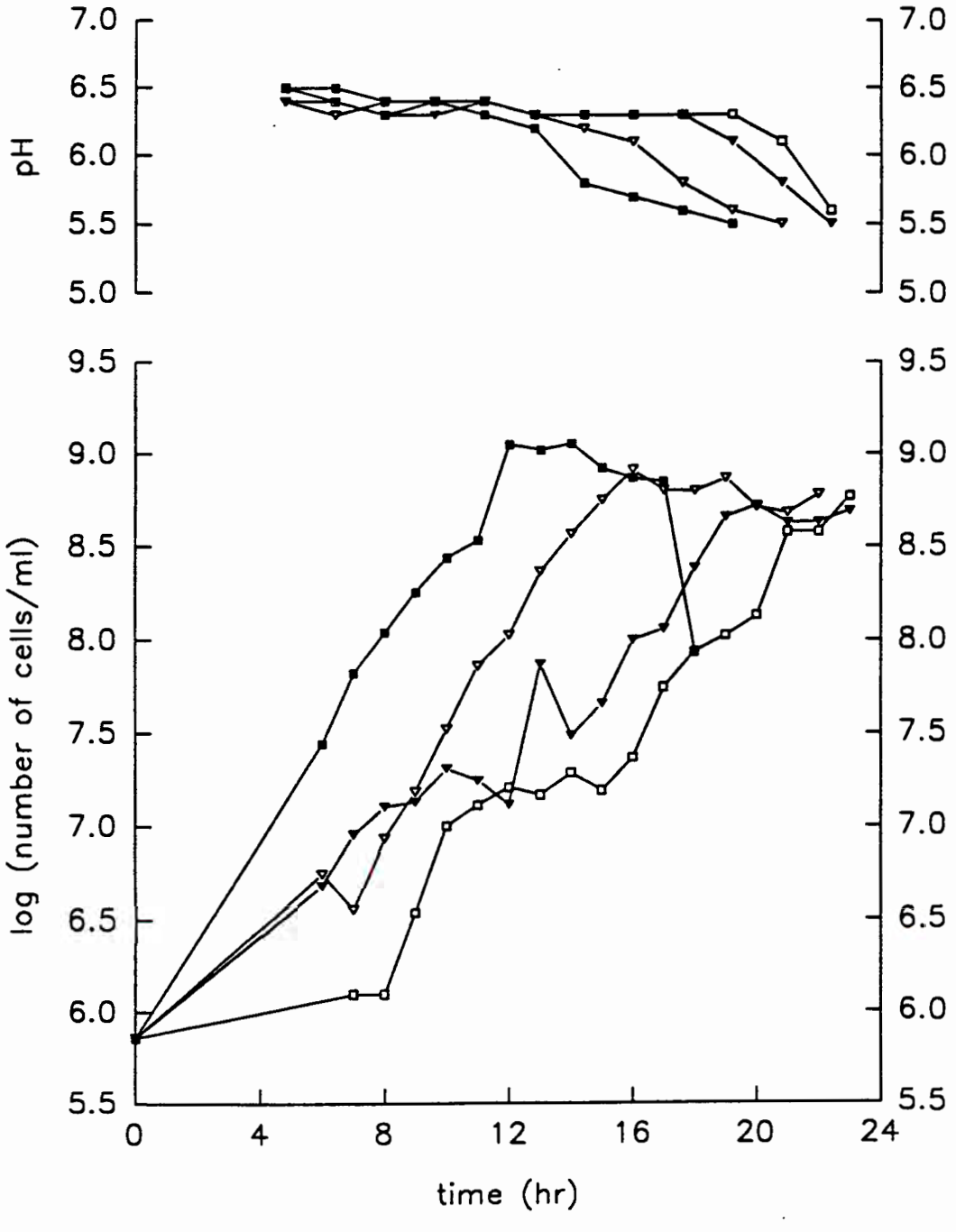


Figure 5.11 Growth curve of *C. longisporum* on salicin (▼), arbutin (▽), cellobiose (■) and glucose (□) medium. The pH was measured every 2 hr. Total cell counts were determined microscopically.

Table 5.11 Comparison of total versus viable cell counts. Samples were taken at the times indicated.

Substrate	Sampling time (hr) after inoculation (refer to Figure 5.11)	Total cell counts log (number of cells/ml)	Viable cell counts log (CFU[a]/ml)
Cellobiose	15	8.92	7.57
Arbutin	18	8.80	7.44
Salicin	21	8.63	7.53
Glucose	23	8.77	7.23

[a]CFU = colony forming units

The variable nature of this organism was demonstrated in later experiments (not shown). Although *C. longisporum* appeared to grow faster on cellobiose and arbutin, than on salicin or glucose (Figure 5.11), these trends tended to be variable. On some occasions, for example, glucose grown cultures grew faster than those growing on arbutin. As the reasons for this variability were not determined, the growth curve presented in Figure 5.11 can therefore only be considered to give a rough estimate of the characteristics of this organism growing on these substrates.

5.4.7 Isolation and analysis of *C. longisporum* RNA

As cultures of *C. longisporum* never reached high cell numbers (highest being 10^7 viable cells/ml), there was insufficient cell mass to isolate RNA during early stages of growth. Furthermore, cell lysis occurred shortly after the cultures reached stationary phase. Thus cultures in mid- to late-log phase (total cell counts 10^{7-8}; viable cell counts 10^{6-7} cells/ml) were used for RNA isolations, and RNA was isolated from cultures grown on cellobiose, arbutin, salicin and glucose (Figure 5.12). RNA was also isolated from *E. coli* containing pUC18 and pBGL1 for use as negative and positive controls, respectively.

CONCLUSIONS

The aim of this thesis was to gain more insight into the mechanisms of lignocellulose degradation in the rumen. Two components were analysed; the cellodextrinase enzyme (CelA) from *R. flavefaciens*, and an aryl-ß-glucoside (*abg*) uptake and utilisation system from *C. longisporum*.

The *abg* system from *C. longisporum* was an important discovery. To our knowledge, it is the first system which links lignocellulose degradation to phenolic glycoside utilisation in the rumen. This system is very similar to the *Enterobacteriaceae* systems, which is also fascinating from an evolutionary viewpoint. Although the *abg* system presented here is an exciting discovery, future analyses are critical to determine the exact role of this system in the host organism.

One possible role for the *abg* system stems from a recent report by Varel *et al.* (1995), where the authors found that *C. longisporum* introduced into the bovine rumen was lost within 48 hr of inoculation. Yet some unknown factors must be contributing to its establishment in the rumen of bison, from which it was originally reisolated (215). As bison can feed on tree-bark (147), containing high levels of phenolic compounds, it is tempting to speculate that the ability of *C. longisporum* to utilise phenolic glycosides (through systems such as the *abg* system) is a contributing factor to its establishment in the rumen.

It is hoped that the analyses of both the cellodextrinase and *abg* system presented in this thesis give a clearer understanding of some of the mechanisms involved in lignocellulose utilisation in the rumen. It is also hoped that these analyses will be beneficial to future research.

REFERENCES

(1) Aiba, H., Adhya, S., and Crombrugghe, B. 1981. Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. **The Journal of Biological Chemistry.** Vol 256(22). pp 11905-11910.

(2) Akin, D.E. 1976. Ultrastructure of rumen bacterial attachment to forage cell walls. **Applied and Environmental Microbiology.** Vol 31(4). pp 562-568.

(3) Akin, D.E. 1980. Evaluation by electron microscopy and anaerobic culture of types of rumen bacteria associated with digestion of forage cell walls. **Applied and Environmental Microbiology.** Vol 39(1). pp 242-252.

(4) Akin, D.E. 1993. Plant structure and chemistry - Influence on lignocellulose degradation by rumen protozoa, bacteria, and fungi. **Genetics, biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. UNI publishers co., Ltd. Tokyo, Japan. pp 95-104.

(5) Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1983. Special features of plant cells. **Molecular biology of the cell.** Garland publishing. New York. Chapter 19. pp 1099-1144.

(6) Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1983. The plasma membrane. **Molecular biology of the cell.** Garland publishing. New York. Chapter 6. pp 256-314.

(7) Alloing, G., Trombe, M.C., and Claverys, J.P. 1990. The *ami* locus of the Gram-positive bacterium *Streptococcus pneumoniae* is similar to binding protein-dependent transport operons of Gram-negative bacteria. **Molecular Microbiology.** Vol 4(4). pp 633-644.

(8) Aminov, R.I., Kaneichi, K., Miyagai, T., and Ohmiya, K. 1993. Conjugal transfer of plasmid pAMβ1 to *Ruminococcus albus*. **Genetics, Biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. UNI publishers. Tokyo, Japan. pp 179-187.

(9) Aminov, R.I., Karita, S., Kobayashi, Y., Kawazu, T., Sakka, K., Hoshino, S., and Shimada, K. 1993. Genetic engineering of cellulolytic rumen bacteria. **Genetics, Biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. UNI publishers co., Ltd. Tokyo, Japan. pp 114-124.

(10) Amster-Choder, O., Houman, F., and Wright, A. 1989. Protein phosphorylation regulates transcription of the β-glucoside utilisation operon in *E. coli*. **Cell.** Vol 58. pp 847-855.

(11) Amster-Choder, O. and Wright, A. 1990. Regulation of activity of a transcriptional antiterminator in *E. coli* by phosphorylation *in vivo*. **Science.** Vol 249. pp 540-542.

(12) Amster-Choder, O. and Wright, A. 1992. Modulation of the dimerisation of a transcriptional antiterminator protein by phosphorylation. **Science.** Vol 257. pp 1395-1398.

(13) Amster-Choder, O. and Wright, A. 1993. Transcriptional regulation of the *bgl* operon of *Escherichia coli* involves phosphotransferase system-mediated phosphorylation of a transcriptional antiterminator. **The Journal of Cell Biology.** Vol 51. pp 83-90.

(14) Annison, E.F. 1984. The metabolism of neutral and acidic lipids by tissues of the ruminant. **Herbivore nutrition in the subtropics and tropics.** Editors: Gilchrist, F.M.C. and Mackie, R.I. The Science Press. Johannesburg, South Africa. Chapter 26. pp 549-571.

(15) Ayers, W.A. 1958. Phosphorylation of cellobiose and glucose by *Ruminococcus flavefaciens*. **Journal of Bacteriology.** Vol 76(5). pp 515-517.

(16) Ayers, W.A. 1959. Phosphorolysis and synthesis of cellobiose by cell extracts from *Ruminococcus flavefaciens*. **The Journal of Biological Chemistry.** Vol 234(11). pp 2819-2822.

(17) Aymerich, S. and Steinmetz, M. 1992. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. **Proceedings of the National Academy of Sciences.** Vol 89. pp 10410-10414.

(18) Bairoch, A. 1994. **Prosite dictionary of protein sites and patterns.** Genetics Computer Group. USA.

(19) Barros, M.E.C. and Thomson, J.A. 1987. Cloning and expression in *Escherichia coli* of a cellulase gene from *Ruminococcus flavefaciens*. **Journal of Bacteriology.** Vol 169(4). pp 1760-1762.

(20) Barstow, D.A., Sharman, A.F., Atkinson, T., and Minton, N.P. 1986. Cloning and complete nucleotide sequence of the *Bacillus stearothermophilus* tryptophanyl tRNA synthetase gene. **Gene.** Vol 46. pp 37-45.

(21) Bayer, E.A., Setter, E., and Lamed, R. 1985. Organisation and distribution of the cellulosome in *Clostridium thermocellum*. **Journal of Bacteriology.** Vol 163(2). pp 552-559.

(22) Beguin, P. 1990. Molecular biology of cellulose degradation. **Annual Review Microbiology.** Vol 44. pp 219-248.

(23) Beguin, P. and Aubert, J. 1994. The biological degradation of cellulose. **FEMS Microbiology Reviews.** Vol 13. pp 25-58.

(24) Berg, B. 1975. Cellulase location in *Cellvibrio fulvus.* **Canadian Journal of Microbiology.** Vol 21. pp 51-57.

(25) Berger, E., Jones, W.A., Jones, D.T., and Woods, D.R. 1989. Cloning and sequencing of an endoglucanase (*end1*) gene from *Butyrivibrio fibrisolvens* H17c. **Molecular and General Genetics.** Vol 219. pp 193-198.

(26) Berger, E., Jones, W.A., Jones, D.T., and Woods, D.R. 1990. Sequencing and expression of a cellodextrinase (*ced1*) gene from *Butyrivibrio fibrisolvens* H17c cloned in *Escherichia coli.* **Molecular and General Genetics.** Vol 223. pp 310-318.

(27) Bernalier, A., Fonty, G., Bonnemoy, F., and Gouet, P. 1993. Inhibition of the cellulolytic activity of *Neocallimastix frontalis* by *Ruminococcus flavefaciens.* **Journal of General Microbiology.** Vol 139. pp 873-880.

(28) Bhat, S., Goodenough, P.W., Owen, E., and Bhat, M.K. 1993. Cellobiose: A true inducer of cellulosome in different strains of *Clostridium thermocellum.* **FEMS Microbiology Letters.** Vol 111. pp 73-78.

(29) Boyd, R.F. 1988. Microbial metabolism and growth. **General Microbiology.** 2nd edition. Editor: Brake, D.K. Time Mirror/Mosby College Publishing. St. Louis, USA. Chapter 12. pp 304-346.

(30) Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry.** Vol 72. pp 248-254.

(31) Bramley, H.F. and Kornberg, H.L. 1987. Nucleotide sequence of *bglC*, the gene specifying Enzyme II^{bgl} of the PEP: sugar phosphotransferase system in *Escherichia coli* K12, and overexpression of the gene product. **Journal of General Microbiology.** Vol 133. pp 563-573.

(32) Brick, P., Bhat, T.N., and Blow, D.M. 1988. Structure of Tyrosyl-tRNA synthetase refined at 2.3 Å resolution. **Journal of Molecular Biology.** Vol 208. pp 83-98.

(33) Brown, G.D., Jorgensen, T., Morris, E.J., and Thomson, J.A. 1993. Analysis of a cellodextrinase cloned from *Ruminococcus flavefaciens* FD-1. **FEMS Microbiology Letters.** Vol 111. pp 57-62.

(34) Bueno, R., Pahel, G., and Magasanik, B. 1985. Role of *glnB* and *glnD* gene products in regulation of the *glnALG* operon of *Escherichia coli*. **Journal of Bacteriology.** Vol 164(2). pp 816-822.

(35) Caldwell, D.R. and Byrant, M.P. 1966. Medium without rumen fluid for nonselective enumeration and isolation of rumen bacteria. **Applied Microbiology.** Vol 14(5). pp 794-801.

(36) Chang, A.C.Y. and Cohen, S.N. 1978. Construction and characterisation of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. **Journal of Bacteriology.** Vol 134(3). pp 1141-1156.

(37) Chesson, A. and Forsberg, C.W. 1988. Polysaccharide degradation by rumen microorganisms. **The Rumen Microbial Ecosystem.** Editor: Hobson, P.N. Elsevier Applied Science. Essex, England. Chapter 8. pp 251-284.

(38) Chow, J.M. and Russell, J.B. 1992. Effect of pH and monoensin on glucose transport by *Fibrobacter succinogenes*, a cellulolytic ruminal bacterium. **Applied and Environmental Microbiology.** Vol 58(4). pp 1115-1120.

(39) Chung, C.T. and Miller, R.H. 1988. A rapid and convenient method for the preparation and storage of competent bacterial cells. **Nucleic Acids Research.** Vol 16(8). pp 3580

(40) Clark, M.F. and Bar-Joseph, M. 1993. Enzyme Immunosorbent Assays in Plant Virology. Chapter 3. pp 51-85.

(41) Colowick, S.P. and Kaplan, N.O. 1976. Multistep enzyme systems. **Methods in Enzymology.** Vol XLIV. pp 457-458.

(42) Cotta, M.A., Wheeler, M.B., and Whitehead, T.R. 1994. Cyclic AMP in ruminal and other bacteria. **FEMS Microbiology Letters.** Vol 124. pp 355-360.

(43) Coutinho, J.B., Gilkes, N.R., Kilburn, D.G., Warren, R.A.J., and Miller, R.C. 1993. The nature of the cellulose-binding domain affects the activities of a bacterial endoglucanase on different forms of cellulose. **FEMS Microbiology Letters.** Vol 113. pp 211-218.

(44) Crutz, A., Steinmetz, M., Aymerich, S., Richter, R., and Coq, D. 1990. Induction of levansucrase in *Bacillus subtilis*: an antitermination mechanism negatively controlled by the phosphotransferase system. **Journal of Bacteriology.** Vol 172(2). pp 1043-1050.

(45) Debarbouille, M., Arnaud, M., Fouet, A., Klier, A., and Rapoport, G. 1990. The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shared strong homology with transcriptional antiterminators. **Journal of Bacteriology.** Vol 172(7). pp 3966-3973.

(46) Demain, A.L. and Lynd, L. 1993. Turning garbage into motor fuel: fanciful dream or feasible scheme?. **Genetics, Biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. UNI publishers. Tokyo, Japan. pp 573-583.

(47) Deshpande, M.V., Eriksson, K., and Pettersson, L.G. 1984. An assay for selective determination of exo-1,4-β-glucanases in a mixture of cellulolytic enzymes. **Analytical Biochemistry.** Vol 138. pp 481-487.

(48) Deutscher, J., Reizer, J., Fischer, C., Galinier, A., Saier, M.H., and Steinmetz, M. 1994. Loss of protein kinase-catalysed phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system, by mutation of the *ptsH* gene confers catabolite repression resistance to several catabolic genes of *Bacillus subtilis*. **Journal of Bacteriology.** Vol 176(11). pp 3336-3344.

(49) Dills, S.S., Apperson, A., Schmidt, M.R., and Saier, M.H. 1980. Carbohydrate transport in bacteria. **Microbiological reviews.** Vol 44(3). pp 385-418.

(50) Dinh, T., Paulsen, I.T., and Saier, M.H. 1994. A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of Gram-negative bacteria. **Journal of Bacteriology.** Vol 176(13). pp 3825-3831.

(51) Doerner, K.C., Howard, G.T., Mackie, R.I., and White, B.A. 1992. β-Glucanase expression by *Ruminococcus flavefaciens* FD-1. **FEMS Microbiology Letters.** Vol 93. pp 147-154.

(52) Eckert, R. and Randall, D. 1983. Feeding, Digestion, and absorption. **Animal Physiology: Mechanisms and adaptions.** Editors: Vapnek, P.C. and Hofmann, L. W.H. Freeman and company. San Francisco, USA. Chapter 15. pp 645-688.

(53) Erni, B. 1992. Group translocation of glucose and other carbohydrates by the bacterial phosphotransferase system. **International review of cytology: a survey of cell biology.** 137th edition. Editors: Bourne, G.H., Jeon, K.W., and Friedlander, M. Academic press. pp 127-148.

(54) Felix, C.R. and Ljungdahl, L.G. 1993. The cellulosome: The exocellular organelle of Clostridium. **Annual Review Microbiology.** Vol 47. pp 791-819.

(55) Ferreira, L.M.A., Durrant, A.J., Hall, J., Hazlewood, G.P., and Gilbert, H.J. 1990. Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. **Biochemistry Journal.** Vol 269. pp 261-264.

(56) Ferreira, L.M.A., Hazlewood, G.P., Barker, P.J., and Gilbert, H.J. 1991. The cellodextrinase from *Pseudomonas fluorescens* subsp. *cellulosa* consists of multiple functional domains. **Biochemistry Journal.** Vol 799. pp 793-799.

(57) Flint, H.J., Martin, J., McPherson, C.A., Daniel, A.S., and Zhang, J. 1993. A bifunctional enzyme, with separate xylanase and β(1,3-1,4)-glucanase domains, encoded by the *xynD* gene of *Ruminococcus flavefaciens*. **Journal of Bacteriology.** Vol 175(10). pp 2943-2951.

(58) Forchhammer, K. and De Marsac, N.T. 1995. Functional analysis of the phosphoprotein P_{II} (*glnB* gene product) in the cyanobacterium *Synechococcus* sp. strain PCC7942. **Journal of Bacteriology.** Vol 177(8). pp 2033-2040.

(59) Forsberg, C.W., Gong, J., Malburg, L.M., Zhu, H., Iyo, A., Chengh, K.J., Krell, P.J., and Phillips, J.P. 1993. Cellulases and hemicellulases of *Fibrobacter succinogenes* and their roles in fibre digestion. **Genetics, Biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. Uni publishers. Tokyo, Japan. pp 125-136.

(60) Fourney, R.M., Miyakoshi, J., Day, R.S., and Paterson, M.C. 1988. Northern blotting: efficient RNA staining and transfer. **Focus.** Vol 10(1). pp 5-6.

(61) Fox, C.F. and Wilson, G. 1968. The role of a phosphoenolpyruvate-dependent kinase system in β-glucoside catabolism in *Escherichia coli.* **Proceedings of the National Academy of Sciences.** Vol 59. pp 988-995.

(62) Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., and Turner, D.H. 1986. Improved free-energy parameters for predictions of RNA duplex stability. **Proceedings of the National Academy of Sciences.** Vol 83. pp 9373-9377.

(63) Fujino, T., Beguin, P., and Aubert, J. 1992. Cloning of a *Clostridium thermocellum* DNA fragment encoding polypeptides that bind the catalytic components of the cellulosome. **FEMS Microbiology Letters.** Vol 94. pp 165-170.

(64) Fujino, T., Beguin, P., and Aubert, J. 1993. Organisation of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. **Journal of Bacteriology.** Vol 175(7). pp 1891-1899.

(65) Gaboriaund, C., Benchetrit, B.T., and Mornon, J.P. 1987. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. **FEBS Letters.** Vol 1. pp 149-155.

(66) Gardner, R.G., Wells, J.E., Russell, J.B., and Wilson, D.B. 1995. The effect of carbohydrates on the expression of the *Prevotella ruminicola* 1,4-β-D-endoglucanase. **FEMS Microbiology Letters.** Vol 125. pp 305-310.

(67) Gardner, R.M., Doerner, K.C., and White, B.A. 1987. Purification and characterization of an exo-β-1,4-glucanase from *Ruminococcus flavefaciens* FD-1. **Journal of Bacteriology.** Vol 169(10). pp 4581-4588.

(68) Genetics Computer Group. 1994. **Program manual for the Wisconsin package, Version 8.0.** USA.

(69) Gerngross, U.T., Romaniec, M.P.M., Kobayashi, T., Huskisson, N.S., and Demain, A.L. 1993. Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal S_1-protein reveals an unusual degree of internal homology. **Molecular Microbiology.** Vol 8(2). pp 325-334.

(70) Ghose, T.K. 1987. Measurement of cellulase activities. **Pure & Applied Chemistry.** Vol 59(2). pp 257-268.

(71) Gilbert, H.J. and Hazlewood, G.P. 1993. Bacterial cellulases and xylanases. **Journal of General Microbiology.** Vol 139. pp 187-194.

(72) Gilkes, N.R., Claeyssens, M., Aebersold, R., Henrissat, B., Meinke, A., Morrison, H.D., Kilburn, D.G., Warren, R.A.J., and Miller, R.C. 1991. Structural and functional relationships in two families of β-1,4-glycanases. **European Journal of Biochemistry.** Vol 202. pp 367-377.

(73) Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., and Warren, R.A.J. 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. **Microbiological reviews.** Vol 55(2). pp 303-315.

(74) Gilson, E., Alloing, G., Schmidt, T., Claverys, J., Dudler, R., and Hofnung, M. 1988. Evidence for high affinity binding-protein dependent transport systems in Gram-positive bacteria and in Mycoplasma. **The EMBO Journal.** Vol 7(12). pp 3971-3974.

(75) Gong, J. and Forsberg, C.W. 1993. Cellulose binding proteins and their potential roles in adhesion of *Fibrobacter succinogenes* to cellulose. **Genetics, Biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. UNI publishers co., Ltd. Tokyo, Japan. pp 138

(76) Gong, J., Lo, R.Y.C., and Forsberg, C.W. 1989. Molecular cloning and expression in *Escherichia coli* of a cellodextrinase gene from *Bacteriodes succinogenes* S85. **Applied and Environmental Microbiology.** Vol 55(1). pp 132-136.

(77) Greenberg, N.M., Warren, R.A.J., Kilburn, D.G., and Miller, R.C. 1987. Regulation and initiation of *cenB* transcripts of *Cellulomonas fimi*. **Journal of Bacteriology.** Vol 169(10). pp 4674-4677.

(78) Grundy, F.J. and Henkin, T.M. 1993. tRNA as a positive regulator of transcription antitermination in *B. subtilis*. **Cell.** Vol 74. pp 475-482.

(79) Grundy, F.J. and Henkin, T.M. 1994. Conservation of a transcriptional antitermination mechanism in amino-acyl-tRNA synthetase and amino acid biosynthesis genes in Gram-positive bacteria. **Journal of Molecular Biology.** Vol 235. pp 798-804.

(80) Hall, B.G. and Faunce, W. 1987. Functional genes for cellobiose utilisation in natural isolates of *Escherichia coli*. **Journal of Bacteriology.** Vol 169(6). pp 2713-2717.

(81) Hall, B.G. and Xu, L. 1992. Nucleotide sequence, function, activation, and evolution of the cryptic *asc* operon of *Escherichia coli* K12. **Molecular Biology and Evolution.** Vol 9. pp 688-706.

(82) Hall, C.V., van Cleemput, M., Muench, K.H., and Yanofsky, C. 1982. The nucleotide sequence of the structural gene for *Escherichia coli* tryptophanyl-tRNA synthetase. **The Journal of Biological Chemistry.** Vol 257(2). pp 6132-6136.

(83) Hall, C.V. and Yanofsky, C. 1981. Cloning and characterization of the gene for *Escherichia coli* tryptophanyl-transfer ribonucleic acid synthetase. **Journal of Bacteriology.** Vol 148(3). pp 941-949.

(84) Hanlon, D.W. and Ordal, G.W. 1994. Cloning and characterisation of genes encoding methyl-accepting chemotaxis proteins in *Bacillus subtilis*. **The Journal of Biological Chemistry.** Vol 269(19). pp 14038-14046.

(85) Hassouni, M.E., Chippaux, M., and Barras, F. 1990. Analysis of the *Erwinia chrysanthemi arb* genes, which mediate metabolism of aromatic β-glucosides. **Journal of Bacteriology.** Vol 172(11). pp 6261-6267.

(86) Hassouni, M.E., Henrissat, B., Chippaux, M., and Barra, F. 1992. Nucleotide sequences of the *arb* genes, which control β-glucoside utilisation in *Erwinia chrysanthemi*: Comparison with the *Escherichia coli bgl* operon and evidence for a new β-glycohydrolase family including enzymes from Eubacteria, Archeabacteria, and humans. **Journal of Bacteriology.** Vol 174(3). pp 765-777.

(87) Helaszek, C.T. and White, B.A. 1991. Cellobiose uptake and metabolism by *Ruminococcus flavefaciens*. **Applied and Environmental Microbiology.** Vol 57(1). pp 64-68.

(88) Hengstenberg, W., Kohlbrecher, D., Wit, E., Kruse, R., Christiansen, I., Peters, D., Strandmann, R.P., Stadtler, P., Koch, B., and Kalbitzer, H. 1993. Structure and function of proteins of the phosphotransferase system and of 6-phospho-β-glycosidases in Gram-positive bacteria. **FEMS Microbiology Reviews.** Vol 12. pp 149-164.

(89) Hengstenberg, W., Reiche, B., Eisermann, R., Fischer, R., Kebler, U., Tarrach, A., De Vos, W.M., Kalbitzer, H., and Glaser, S. 1989. Structure and function of proteins involved in sugar transport by the PTS of Gram-positive bacteria. **FEMS Microbiology Reviews.** Vol 63. pp 35-42.

(90) Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. **Gene.** Vol 28. pp 351-359.

(91) Henkin, T.M. 1994. tRNA-directed transcription antitermination. **Molecular Microbiology.** Vol 13(3). pp 381-387.

(92) Henkin, T.M., Glass, B.L., and Grundy, F.J. 1992. Analysis of the *Bacillus subtilis tyrS* gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. **Journal of Bacteriology.** Vol 174(4). pp 1299-1306.

(93) Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochemistry Journal.** Vol 280. pp 309-316.

(94) Henrissat, B. and Bairoch, A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. **Biochemistry Journal.** Vol 293. pp 781-788.

(95) Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J.-P. 1989. Cellulase families revealed by hydrophobic cluster analysis. **Gene.** Vol 81. pp 83-95.

(96) Henrissat, B., Driguez, H., Viet, C., and Schulein, M. 1985. Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. **Bio/Technology.** Vol 3. pp 722-726.

(97) Higgnins, D.G. and Sharp, P.M. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. **Gene.** Vol 73. pp 237-244.

(98) Hiltner, P. and Dehority, B.A. 1983. Effect of soluble carbohydrates on digestion of cellulose by pure cultures of rumen bacteria. **Applied and Environmental Microbiology.** Vol 46(3). pp 642-648.

(99) Hoischen, C., Dijkstra, A., Rottem, S., Reizer, J., and Saier, M.H. 1993. Presence of protein constituents of the Gram-positive bacterial phosphotransferase regulatory system in *Acholeplasma laidlawii.* **Journal of Bacteriology.** Vol 175(20). pp 6599-6604.

(100) Holt, J.G., Sneath, P.H.A., Mair, N.S., and Sharpe, E.M. 1984. **Bergey's manual of systematic bacteriology.** Williams and Wilkins. Baltimore, USA.

(101) Holtel, A. and Merrick, M. 1988. Identification of the *Klebsiella pneumoniae glnB* gene: nucleotide sequence of wild-type and mutant alleles. **Molecular and General Genetics.** Vol 215. pp 134-138.

(102) Houman, F., Diaz-Torres, M.R., and Wright, A. 1990. Transcriptional antitermination in the *bgl* operon of *E. coli* is modulated by a specific RNA binding protein. **Cell.** Vol 62. pp 1153-1163.

(103) Howard, G.T. and White, B.A. 1988. Molecular cloning and expression of cellulase genes from *Ruminococcus albus* 8 in *Escherichia coli* bacteriophage lambda. **Applied and Environmental Microbiology.** Vol 54(7). pp 1752-1755.

(104) Huang, C., Kelly, W.J., Asmundson, R.V., and Yu, P. 1989. Molecular cloning and expression of multiple cellulase genes from *Ruminococcus flavefaciens* strain 186 in *Escherichia coli.* **Applied Microbiological Biotechnology.** Vol 31. pp 265-271.

(105) Huang, L. and Forsberg, C.W. 1987. Isolation of a cellodextrinase from *Bacteroides succinogenes.* **Applied and Environmental Microbiology.** Vol 53(5). pp 1034-1041.

(106) Huang, L. and Forsberg, C.W. 1988. Purification and comparison of the periplasmic and extracellular forms of the cellodextrinase from *Bacteroides succinogenes.* **Applied and Environmental Microbiology.** Vol 54(6). pp 1488-1493.

(107) Huang, L. and Forsberg, L. 1990. Cellulose digestion and cellulase regulation and distribution in *Fibrobacter succinogenes* subsp. *succinogenes* S85. **Applied and Environmental Microbiology.** Vol 56(5). pp 1221-1228.

(108) Hueck, C.J. and Hillen, W. 1995. Catabolite repression in *Bacillus subtilis*: a global regulatory mechanism for the Gram-positive bacteria?. **Molecular Microbiology.** Vol 15(3). pp 395-401.

(109) Hungate, R.E. 1988. Introduction: The ruminant and the rumen. **The Rumen Microbial Ecosystem.** Editor: Hobson, P.N. Elsevier Applied science. Essex, England. Chapter 1. pp 1-20.

(110) Izard, J.W. and Kendall, D.A. 1994. Signal peptides: exquisitely designed transport promoters. **Molecular Microbiology.** Vol 13(5). pp 765-773.

(111) Jones, D.H., Franklin, C.H., and Thomas, C.M. 1994. Molecular analysis of the operon which encodes the RNA polymerase sigma factor δ^{54} of *Escherichia coli*. **Microbiology.** Vol 140. pp 1035-1043.

(112) Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.J., Claeyssens, M., Beguin, P., and Aubert, J. 1992. Three-dimensional structure of a thermostable bacterial cellulase. **Nature.** Vol 357. pp 89-91.

(113) Knowles, J., Teeri, T.T., Lehtovaara, P., Penttila, M., and Saloheimo, M. 1988. The use of gene technology to investigate fungal cellulolytic enzymes. **FEMS Symposium: Biochemistry and genetics of cellulose degradation.** 43th edition. Editors: Aubert, J., Beguin, P., and Millet, J. Academic Press. London, England. pp 153-170.

(114) Kraulis, P.J., Clore, G.M., Nilges, M.J., Jones, T.A., Pettersson, G., Knowles, J., and Gronenborn, A.M. 1989. Determination of the three-dimensional structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. **Biochemistry.** Vol 28. pp 7241-7257.

(115) Kruger, S., Stulke, J., and Hecker, M. 1993. Catabolite repression of β-glucanase synthesis in *Bacillus subtilis*. **Journal of General Microbiology.** Vol 139. pp 2047-2054.

(116) Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydrophobic character of a protein. **Journal of Molecular Biology.** Vol 157. pp 105-132.

(117) Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature.** Vol 227. pp 680-685.

(118) Lai, X. and Ingram, L.O. 1993. Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. **Journal of Bacteriology.** Vol 175(20). pp 6441-6450.

(119) Leatherwood, J.M. 1973. Cellulose degradation by *Ruminococcus*. **Federation Proceedings.** Vol 32(7). pp 1814-1818.

(120) Lin, L., Rumbak, E., Zappe, H., Thomson, J.A., and Woods, D.R. 1990. Cloning, sequencing and analysis of expression of a *Butyrivibrio fibrisolvens* gene encoding a β-glucosidase. **Journal of General Microbiology.** Vol 136. pp 1567-1576.

(121) Lineweaver, H. and Burk, D. 1934. The determination of enzyme dissociation constants. **Journal of the American Chemical Society.** Vol 56. pp 658-666.

(122) Ljungdahl, L.G., Coughlan, M.P., Mayer, F., Mori, Y., Hon-Nami, H., and Hon-Nami, K. 1988. Macrocellulase complexes and yellow affinity substance from *Clostridium thermocellum*. **Methods in Enzymology.** Vol 160. pp 483-500.

(123) Lo, T.C.Y., Rayman, M.K., and Sanwal, B.D. 1972. Transport of succinate in *Escherichia coli*. **The Journal of Biological Chemistry.** Vol 247(19). pp 6323-6331.

(124) Lukat, G.S. and Stock, J.B. 1993. Response regulation in bacterial chemotaxis. **Journal of Cellular Biochemistry.** Vol 51. pp 41-46.

(125) Magasanik, B. 1993. The regulation of nitrogen utilisation in enteric bacteria. **Journal of Cellular Biochemistry.** Vol 51. pp 34-40.

(126) Mahadevan, S., Reynolds, A.E., and Wright, A. 1987. Positive and negative regulation of the *bgl* operon in *Escherichia coli*. **Journal of Bacteriology.** Vol 169(6). pp 2570-2578.

(127) Mahadevan, S. and Wright, A. 1987. A bacterial gene involved in transcription antitermination: regulation at a Rho-independent terminator in the *bgl* operon of *E. coli*. **Cell.** Vol 50. pp 485-494.

(128) Malburg, L.M. and Forsberg, C.W. 1993. *Fibrobacter succinogenes* S85 possesses at least nine distinct glucanase genes. **Canadian Journal of Microbiology.** Vol 39. pp 882-891.

(129) Martin, S.A. and Russell, J.B. 1986. Phosphoenolpyruvate-dependent phosphorylation of hexoses by ruminal bacteria: evidence for the phosphotransferase transport system. **Applied and Environmental Microbiology.** Vol 52(6). pp 1348-1352.

(130) Martin, S.A. and Russell, J.B. 1987. Transport and phosphorylation of disaccharides by the ruminal bacterium *Streptococcus bovis*. **Applied and Environmental Microbiology.** Vol 53(10). pp 2388-2393.

(131) Martin, S.A. and Russell, J.B. 1988. Mechanisms of sugar transport in the rumen bacterium *Selenomonas ruminantium.* **Journal of General Microbiology.** Vol 134. pp 819-827.

(132) McGavin, M., Lam, J., and Forsberg, C.W. 1990. Regulation and distribution of *Fibrobacter succinogenes* subsp. *succinogenes* S85 endoglucanases. **Applied and Environmental Microbiology.** Vol 56(5). pp 1235-1244.

(133) Meadow, N.D., Fox, D.K., and Roseman, S. 1990. The bacterial phosphoenolpyruvate: glycose phosphotransferase system. **Annual Review Biochemistry.** Vol 59. pp 497-542.

(134) Messing, J., Crea, R., and Seeburg, P.H. 1981. A system for shotgun DNA sequencing. **Nucleic Acids Research.** Vol 9(2). pp 309-321.

(135) Messner, R., Kubicek-Pranz, E.M., Gsur, A., and Kubicek, C.P. 1993. Cellobiohydrolase II is the main conidial-bound cellulase in *Trichoderma reesei* and other *Trichoderma* strains. **Archives of Microbiology.** Vol 155. pp 601-606.

(136) Millward-Sadler, S.J., Poole, D.M., Henrissat, B., Hazlewood, G.P., Clarke, J.H., and Gilbert, H.J. 1993. Molecular dissection of modular xylanase from *Cellulomonas fimi.* **Genetics, Biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. UNI publishers co., Ltd. Tokyo, Japan. pp 298-307.

(137) Mittendorf, V. and Thomson, J.A. 1993. Cloning of an endo-(1-4)-β-glucanase gene, *celA,* from the rumen bacterium *Clostridium* sp. ('*C. longisporum*') and characterization of its product, CelA, in *Escherichia coli.* **Journal of General Microbiology.** Vol 139. pp 3233-3242.

(138) Morse, H.G., Penberthy, W.K., and Morse, M.L. 1971. Biochemical characterisation of the *ctr* mutants of *Escherichia coli*. **Journal of Bacteriology.** Vol 108(2). pp 690-694.

(139) Murphy, N., McConnell, D.J., and Cantwell, B.A. 1984. The DNA sequence of the gene and genetic control sites for the excreted *B. subtilis* enzyme β-glucanase. **Nucleic Acids Research.** Vol 12(13). pp 5355-5367.

(140) Nagel, G.M. and Doolittle, R.F. 1991. Evolution and relatedness in two aminoacyl-tRNA synthetase families. **Proceedings of the National Academy of Sciences.** Vol 88. pp 8121-8125.

(141) National Centre For Biotechnology Information. 1994. **BLAST.** USA.

(142) Nikaido, H. 1992. Porins and specific channels of bacterial outer membranes. **Molecular Microbiology.** Vol 6(4). pp 435-442.

(143) Nikaido, H. and Saier, M.H. 1992. Transport proteins in bacteria: common themes in their design. **Science.** Vol 258. pp 936-942.

(144) Ninfa, A.J. and Magasanik, B. 1986. Covalent modification of the *glnG* product, NRI, by the *glnL* product, NRII, regulates the transcription of the *glnALG* operon in *Escherichia coli*. **Proceedings of the National Academy of Sciences.** Vol 83. pp 5909-5913.

(145) Odenyo, A.A., Mackie, R.I., Sthal, D.A., and White, B.A. 1994. The use of 16S rRNA-targeted oligonucleotide probes to study competition between ruminal fibrolytic bacteria: development of probes for *Ruminococcus* species and evidence for bacteriocin production. **Applied and Environmental Microbiology.** Vol 60(10). pp 3688-3696.

(146) Ollis, D. and White, S. 1990. Protein Crystallization. **Methods in Enzymology.** Vol 182. pp 646-659.

(147) Online Computer Systems, Inc. 1993. **The New Grolier: multimedia encyclopedia.** Grolier Inc.

(148) Onodera, R. 1993. Interrelationships between bacteria, protozoa and fungi in fibre digestion, and skataole and tryptophan production in the rumen ecosystem. **Genetics, Biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. UNI publishers co., Ltd. Tokyo, Japan. pp 105-113.

(149) Orpin, C.G. and Joblin, K.N. 1988. The rumen anaerobic fungi. **The Rumen Microbial Ecosystem.** Editor: Hobson, P.N. Elsevier Applied Science. Essex, England. Chapter 4. pp 129-150.

(150) Pardee, A.B., Jacob, F., and Monod, J. 1959. The genetic control and cytoplasmic expression of "Inducibility" in the synthesis of β-galactosidase by *E. coli.* **Journal of Molecular Biology.** Vol 1. pp 165-178.

(151) Parker, L.L. and Hall, B.G. 1990. Characterisation and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. **Genetics.** Vol 124. pp 455-471.

(152) Peterkofsky, A., Reizer, A., Reizer, J., Gollop, N., Zhu, P., and Amin, N. 1993. Bacterial adenylyl cyclases. **Progress in Nucleic Acid Research and Molecular Biology.** Vol 44. pp 31-65.

(153) Pettipher, G.L. and Latham, M.J. 1979. Characteristics of enzymes produced by *Ruminococcus flavefaciens* which degrade plant cell walls. **Journal of General Microbiology.** Vol 110. pp 21-27.

(154) Poole, D.M., Hazlewood, G.P., Laurie, J.I., Barker, P.J., and Gilbert, H.J. 1990. Nucleotide sequence of the *Ruminococcus albus* SY3 endoglucanase genes *celA* and *celB*. **Molecular and General Genetics.** Vol 223. pp 217-223.

(155) Postma, P.W. and Lengeler, J.W. 1985. Phosphoenolpyruvate: carbohydrate phosphotransferase system of bacteria. **Microbiological reviews.** Vol 49(3). pp 232-269.

(156) Postma, P.W., Lengeler, J.W., and Jacobson, G.R. 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. **Microbiological reviews.** Vol 57(3). pp 543-594.

(157) Ramseier, T.M., Negre, D., Cortay, J.C., Scarabel, M., Cozzone, A.J., and Saier, M.H. 1993. *In vitro* binding of the pleiotropic transcriptional regulatory protein, FruR, to the *fru, pps, ace, pts* and *icd* operons of *Escherichia coli* and *Salmonella typhimurium*. **Journal of Molecular Biology.** Vol 234. pp 28-44.

(158) Rasmussen, M.A., Hespell, R.B., White, B.A., and Bothast, R.J. 1988. Inhibitory effects of methylcellulose on cellulose degradation by *Ruminococcus flavefaciens*. **Applied and Environmental Microbiology.** Vol 54(4). pp 890-897.

(159) Reizer, J., Romano, A.H., and Deutscher, J. 1993. The role of phosphorylation of HPr, a phosphocarrier protein of the phosphotransferase system, in the regulation of carbon metabolism in Gram-positive bacteria. **Journal of Cellular Biochemistry.** Vol 51. pp 19-24.

(160) Reverchon, S., Nasser, W., and Robert-Baudouy, J. 1994. *pecS*: a locus controlling pectinase, cellulase and blue pigment production in *Erwinia chrysanthemi*. **Molecular Microbiology.** Vol 11(6). pp 1127-1139.

(161) Reynolds, A.E., Felton, J., and Wright, A. 1981. Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K12. **Nature.** Vol 293. pp 625-629.

(162) Riverside Scientific. 1993. **Genepro Version 5.00.**

(163) Robyt, J.F. and White, B.J. 1987. Qualitative and quantitative methods for determining biological molecules. **Biochemical Techniques: Theory and Practice.** Editor: Griffing, B. Brooks/Cole Publishing Company. California, USA. Chapter 7. pp 247

(164) Roger, V., Fonty, G., Komisarczuk-Bony, S., and Gouet, P. 1990. Effects of physicochemical factors on the adhesion to cellulose Avicel of the ruminal bacteria *Ruminococcus flavefaciens* and *Fibrobacter succinogenes* subsp. *succinogenes*. **Applied and Environmental Microbiology.** Vol 56(10). pp 3081-3087.

(165) Rosenberg, M. and Court, D. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. **Annual Review of Genetics.** Vol 13. pp 19-53.

(166) Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C., and Jones, T.A. 1990. The three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. **Science.** Vol 249. pp 380-386.

(167) Russell, J.B. 1984. Factors influencing competition and composition of the rumen bacterial flora. **Herbivore nutrition in the tropics and subtropics.** Editors: Gilchrist, F.M.C. and Mackie, R.I. The Science Press. Johannesburg, South Africa. Chapter 15. pp 313-345.

(168) Russell, J.B. and Baldwin, R.L. 1978. Substrate preferences in rumen bacteria: evidence of catabolite regulatory mechanisms. **Applied and Environmental Microbiology.** Vol 36(2). pp 319-329.

(169) Russell, J.B. and Wallace, R.J. 1988. Energy yielding and consuming reactions. **The Rumen Microbial Ecosystem.** Editor: Hobson, P.N. Elsevier applied science. Essex, England. Chapter 6. pp 185-216.

(170) Russell, R.R.B., Aduse-Opoku, J., Sutcliffe, I.C., Tao, L., and Ferretti, J.J. 1992. A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. **The Journal of Biological Chemistry.** Vol 267(7). pp 4631-4637.

(171) Sacerdot, C., Fayat, G., Dessen, P., Springer, M., Plumbridge, J.A., Grunberg-Manago, M., and Blanquet, S. 1982. Sequence of a 1.26-kb DNA fragment containing the structural gene for *E. coli* initiation factor IF3: presence of an AUU initiator codon. **The EMBO Journal.** Vol 1(3). pp 311-315.

(172) Saier, M.H. 1989. Protein phosphorylation and allosteric control of inducer exclusion and catabolite repression by the bacterial phosphoenolpyruvate: sugar phosphotransferase system. **Microbiological reviews.** Vol 53(1). pp 109-120.

(173) Saier, M.H. 1991. A multiplicity of potential carbon catabolite repression mechanisms in prokaryotic and eukaryotic microorganisms. **The New Biologist.** Vol 3(12). pp 1137-1147.

(174) Saier, M.H. 1993. Regulatory interactions involving the proteins of the phosphotransferase system in enteric bacteria. **Journal of Cellular Biochemistry.** Vol 51. pp 62-68.

(175) Saier, M.H. and Reizer, J. 1992. Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. **Journal of Bacteriology.** Vol 174(5). pp 1433-1438.

(176) Saier, M.H., Simoni, R.D., and Roseman, S. 1976. Sugar transport: properties of mutant bacteria defective in proteins of the phosphoenolpyruvate sugar phosphotransferase system. **The Journal of Biological Chemistry.** Vol 251(21). pp 6584-6597.

(177) Saier, M.H., Werner, P.K., and Muller, M. 1989. Insertion of proteins into bacterial membranes: mechanisms, characteristics, and comparisons with the eukaryotic process. **Microbiological reviews.** Vol 53(3). pp 333-366.

(178) Saier, M.H., Ye, J.J., Ramseier, T.M., Figge, R.M., Bledig, S.A., Hueck, C., Deutscher, J., Hoischen, C., Pitaknarongphorn, S., Tetgemeyer, F., Mason, R., Reizer, A., Vartak, N., Michotey, V., Cui, X., and Reizer, J. 1994. The bacterial phosphotransferase system: A multifaceted regulatory system controlling carbon and energy metabolism. **Molecular biology of phosphate metabolism in microorganisms.** Editors: Torriani, A.M., Silver, S., and Yagil, E. ASM Press. Washington, D.C.

(179) Salamitou, S., Lemaire, M., Fujino, T., Ohayon, H., Gounon, P., Beguin, P., and Aubert, J. 1994. Subcellular localisation of *Clostridium thermocelum* ORF3p, a protein carrying a receptor for the docking sequence borne by the catalytic components of the cellulosome. **Journal of Bacteriology.** Vol 176(10). pp 2828-2834.

(180) Salamitou, S., Raynaud, O., Lemaire, M., Coughlan, M., Beguin, P., and Aubert, J. 1994. Recognition of the duplicated segments present in *Clostridium thermocellum* endoglucanase CelD and in the cellulosome-integrating protein CipA. **Journal of Bacteriology.** Vol 176(10). pp 2822-2827.

(181) Salamitou, S., Tokatlidis, K., Beguin, P., and Aubert, J. 1992. Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. **FEBS Letters.** Vol 304(1). pp 89-92.

(182) Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. **Molecular Cloning A laboratory manual.** 2nd edition. Cold Spring Harbor Laboratory Press. New York.

(183) Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences.** Vol 74(12). pp 5463-5467.

(184) Sato, T., Daimon, K., and Sakaguchi, H. 1993. Application of cellulases for textile industry. **Genetics, Biochemistry and ecology of lignocellulose degradation.** Editors: Shimada, K., Ohmiya, K., Kobayashi, Y., Hoshino, S., Sakka, K., and Karita, S. UNI publishers. Tokyo, Japan. pp 602-603.

(185) Schaefler, S. and Malamy, A. 1969. Taxonomic investigations on expressed and cryptic phospho-β-glucosidases in *Enterobacteriaceae*. **Journal of Bacteriology.** Vol 99(2). pp 422-433.

(186) Schimmel, P. 1987. Aminoacyl tRNA synthetases: general scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. **Annual review of Biochemistry.** Vol 56. pp 125-158.

(187) Schimmel, P. 1991. Classes of aminoacyl-tRNA synthetases and the establishment of the genetic code. **Trends in Biochemical Sciences.** Vol 16. pp 1-3.

(188) Schnetz, K. and Rak, B. 1988. Regulation of the *bgl* operon of *Escherichia coli* by transcriptional antitermination. **The EMBO Journal.** Vol 7(10). pp 3271-3277.

(189) Schnetz, K. and Rak, B. 1990. β-Glucoside permease represses the *bgl* operon of *Escherichia coli* by phosphorylation of the antiterminator protein and also interacts with glucose-specific enzyme III, the key element in catabolite control. **Proceedings of the National Academy of Sciences.** Vol 87. pp 5074-5078.

(190) Schnetz, K. and Rak, B. 1992. IS5: A mobile enhancer of transcription in *Escherichia coli*. **Proceedings of the National Academy of Sciences.** Vol 89. pp 1244-1248.

(191) Schnetz, K., Sutrina, S.L., Saier, M.H., and Rak, B. 1990. Identification of catalytic residues in the β-glucoside permease of *Escherichia coli* by site-specific mutagenesis and demonstration of interdomain cross-reactivity between the β-glucoside and glucose systems. **The Journal of Biological Chemistry.** Vol 265(23). pp 13464-13471.

(192) Schnetz, K., Toloczuki, C., and Rak, B. 1987. β-Glucoside (*bgl*) operon of *Escherichia coli* K-12: Nucleotide sequence, genetic organisation, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. **Journal of Bacteriology.** Vol 169(6). pp 2579-2590.

(193) Shine, J. and Dalgarno, L. 1974. The 3'-terminal sequence of the *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. **Proceedings of the National Academy of Sciences.** Vol 71(4). pp 1342-1346.

(194) Shoemaker, S., Watt, K., Tsitovsky, G., and Cox, R. 1983. Characterisation and properties of cellulases purified from *Trichoderma reesei* strain L27. **Bio/Technology.** pp 687-690.

(195) Singh, R.N. and Akimenko, V.K. 1994. Synergism among three purified cellulolytic components of *Clostridium thermocellum.* **FEMS Microbiology Letters.** Vol 122. pp 257-262.

(196) Sinnot, M.L. 1990. Catalytic mechanisms of enzymatic glycosyl transfer. **Chemical Reviews.** Vol 90. pp 1171-1202.

(197) Son, H.S. and Rhee, S.G. 1987. Cascade control of *Escherichia coli* glutamine synthetase. **The Journal of Biological Chemistry.** Vol 262(18). pp 8690-8695.

(198) Spezio, M., Wilson, D.B., and Karplus, P.A. 1993. Crystal structure of the catalytic domain of a thermophilic endocellulase. **Biochemistry.** Vol 32. pp 9906-9916.

(199) Srisodsuk, M., Reinikainen, T., Penttila, M., and Teeri, T.T. 1993. Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. **The Journal of Biological Chemistry.** Vol 268(28). pp 20756-20761.

(200) Stewart, C.S. and Bryant, M.P. 1988. The rumen bacteria. **The Rumen Microbial Ecosystem.** Editor: Hobson, P.N. Elsevier Applied Science. Essex, England. Chapter 2. pp 21-76.

(201) Stewart, C.S., Duncan, S.H., and Flint, H.J. 1990. The properties and forms of *Ruminococcus flavefaciens* which differ in their ability to degrade cotton cellulose. **FEMS Microbiology Letters.** Vol 72. pp 47-50.

(202) Stewart, G.C. 1993. Catabolite repression in the Gram-positive bacteria: generation of negative regulators of transcription. **Journal of Cellular Biochemistry.** Vol 51. pp 25-28.

(203) Strobel, H.J. 1993. Evidence for catabolite inhibition in regulation of pentose utilisation and transport in the ruminal bacterium *Selenomonas ruminantium.* **Applied and Environmental Microbiology.** Vol 59(1). pp 40-46.

(204) Stuelke, J., Schnetz, K., Krieg, M., Krueger, S., Hecker, M., and Rak, B. A transcriptional antiterminator of the BglG family positively regulates utilization of beta-glucans by *Bacillus subtilis.* **Unpublished.**

(205) Stutzenberger, F. 1985. Regulation of cellulolytic activity. **Annual Reports on Fermentation processes.** 8th edition. Academic press. Chapter 6. pp 111-154.

(206) Tam, R. and Saier, M.H. 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. **Microbiological reviews.** Vol 57(2). pp 320-346.

(207) Teather, R.M. and Wood, P.J. 1982. Use of congo red-polysaccharide interactions in the enumeration and characterisation of cellulolytic bacteria from the bovine rumen. **Applied and Environmental Microbiology.** Vol 43(4). pp 777-780.

(208) Thomson, J.A. 1993. Molecular biology of xylan degradation. **FEMS Microbiology Reviews.** Vol 104. pp 65-82.

(209) Thurston, B., Dawson, K.A., and Strobel, H.J. 1993. Cellobiose versus glucose utilisation by the ruminal bacterium *Ruminococcus albus.* **Applied and Environmental Microbiology.** Vol 59(8). pp 2631-2637.

(210) Thurston, B., Dawson, K.A., and Strobel, H.J. 1994. Pentose utilisation by the ruminal bacterium *Ruminococcus albus*. **Applied and Environmental Microbiology.** Vol 60(4). pp 1087-1092.

(211) Titgemeyer, F. 1993. Signal transduction in chemotaxis mediated by the bacterial phosphotransferase system. **Journal of Cellular Biochemistry.** Vol 51. pp 69-74.

(212) Tsinoremas, N.F., Castets, A.M., Harrison, M.A., Allen, J.F., and De Marsac, N.T. 1991. Photosynthetic electron transport controls nitrogen assimilation in cyanobacteria by means of posttranslational modification of the *glnB* gene product. **Proceedings of the National Academy of Sciences.** Vol 88. pp 4565-4569.

(213) Usdin, K., Zappe, H., Jones, D.T., and Woods, D.R. 1986. Cloning, expression, and purification of glutamine synthetase from *Clostridium acetobutylicum*. **Applied and Environmental Microbiology.** Vol 52(3). pp 413-419.

(214) Van Tilbeurgh, H. and Claeyssens, M. 1985. Detection and differentiation of cellulase components using low molecular mass fluorogenic substrates. **FEBS Letters.** Vol 187(2). pp 283-288.

(215) Varel, V.H. 1989. Reisolation and characterization of *Clostridium longisporum*, a ruminal sporeforming cellulolytic anaerobe. **Archives of Microbiology.** Vol 152. pp 209-214.

(216) Varel, V.H. and Pond, W.G. 1992. Characteristics of a new cellulolytic *Clostridium* sp. isolated from pig intestinal tract. **Applied and Environmental Microbiology.** Vol 58(5). pp 1645-1649.

(217) Varel, V.H., Richardson, A.J., and Stewart, C.S. 1989. Degradation of barley straw, rye grass, and alfalfa cell walls by *Clostridium longisporum* and *Ruminococcus albus*. **Applied and Environmental Microbiology.** Vol 55(12). pp 3080-3084.

(218) Varel, V.H., Yen, J.T., and Kreikemeier, K.K. 1995. Addition of cellulolytic *Clostridia* to the bovine rumen and pig intestinal tract. **Applied and Environmental Microbiology.** Vol 61(3). pp 1116-1119.

(219) Wang, W. and Thomson, J.A. 1990. Nucleotide sequence of the *celA* gene encoding a cellodextrinase of *Ruminococcus flavefaciens* FD-1. **Molecular and General Genetics.** Vol 222. pp 265-269.

(220) Wang, W. and Thomson, J.A. 1992. Nucleotide sequence of the *celA* gene encoding a cellodextrinase of *Ruminococcus flavefaciens* FD-1. (Erratum) **Molecular and General Genetics.** Vol 233. pp 492

(221) Webster, T., Tsai, H., Kula, M., Mackie, G.A., and Schimmel, P. 1984. Specific sequence homology and three dimensional structure of an aminoacyl transfer RNA synthetase. **Science.** Vol 226. pp 1315-1317.

(222) Williams, A.G. and Coleman, G.S. 1988. The rumen protozoa. **The Rumen Microbial Ecosystem.** Editor: Hobson, P.N. Elsevier Applied Science. Essex, England. Chapter 3. pp 77-128.

(223) Williams, A.G. and Withers, S.E. 1992. The regulation of xylanolytic enzyme formation by *Butyrivibrio fibrisolvens* NCFB 2249. **Letters in Applied Microbiology.** Vol 14. pp 194-198.

(224) Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hiller, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laisster, N., Latreille, P., Lightning, J., Lloyd, C., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Meig, J., Thomas, K., Vaudin, M., Vaughan, K., Waterson, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohldman, P. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. **Nature.** Vol 368. pp 32-38.

(225) Withers, S.G., Warren, R.A.J., Street, I.P., Ruptiz, K., Kempton, J.B., and Aerbersold, R. 1990. Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a "retaining" glycosidase. **Journal of the American Chemical Society.** Vol 112. pp 5887-5889.

(226) Wolin, M.J. and Miller, T.L. 1988. Microbe-Microbe interactions. **The rumen microbial ecosystem.** Editor: Hobson, P.N. Elsevier applied science. London, England. Chapter 11. pp 343-360.

(227) Wood, T.M. 1988. Preparation of crystalline, amorphous, and dyed cellulase substrates. **Methods in Enzymology.** Vol 160. pp 19-25.

(228) Wood, W.B. 1966. Host specificity of DNA produced by *Escherichia coli*: Bacterial mutations affecting restriction and modification of DNA. **Journal of Molecular Biology.** Vol 16. pp 118-133.

(229) Wray, L.V., Atkinson, M.R., and Fisher, S.H. 1994. The nitrogen-regulation *Bacillus subtilis nrgAB* operon encodes a membrane protein and a protein highly similar to the *Escherichia coli glnB*-encoded PII protein. **Journal of Bacteriology.** Vol 176(1). pp 108-114.

(230) Yanisch-Perron, C., Vieira, J., and Messing, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. **Gene.** Vol 33. pp 103-119.

(231) Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. **Nature.** Vol 289. pp 751-758.

(232) Ye, J.J., Neal, J.W., Cui, X., Reizer, J., and Saier, M.H. 1994. Regulation of the glucose: H^+ symporter by metabolite-activated ATP-dependent phosphorylation of HPr in *Lactobacillus brevis.* **Journal of Bacteriology.** Vol 176(12). pp 3484-3492.

(233) Yoon, K.-H. Cloning and nucleotide sequence of enzyme II of *Brevibacterium lactofermentum* phosphotransferase system. **Unpublished.**

(234) Zabeau, M. and Stanley, K.K. 1982. Enhanced expression of cro-β-galactosidase fusion proteins under the control of the Pr promoter of bacteriophage lambda. **The EMBO Journal.** Vol 1(10). pp 1217-1224.

(235) Zamaroczy, M., Delorme, F., and Elmerich, C. 1990. Characterisation of three different nitrogen-regulated promoter regions for the expression of *glnB* and *glnA* in *Azospirillum brasilense.* **Molecular and General Genetics.** Vol 224. pp 421-430.

(236) Zehnder, A.J.B. 1988. Anaerobic microbial degradation of lignocellulose. **Biology of anaerobic microorganisms.** John Wiley and sons. USA. Chapter 7. pp 333-372.

(237) Zuker, M. 1989. Computer prediction of RNA structure. **Methods in Enzymology.** Vol 180. pp 262-288.

(238) Zuker, M. and Stiegler, P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxilliary information. **Nucleic Acids Research.** Vol 9(1). pp 133-148.

(239) Zukowski, M.M., Miller, L., Cogswell, P., Chen, K., Aymerich, S., and Steinmetz, M. 1990. Nucleotide sequence of the *sacS* locus of *Bacillus subtilis* reveals the presence of two regulatory genes. **Gene.** Vol 90. pp 153-155.